LICENCIATURA EM MATEMÁTICA

DEPARTAMENTO DE FÍSICA, QUÍMICA E MATEMÁTICA

REGRESSÃO LINEAR E NÃO LINEAR SIMPLES APLICADA À ECONOMIA E FINANÇAS

DELSA KOBATA DOS SANTOS LAHR

Sorocaba

2017

UNIVERSIDADE FEDERAL DE SÃO CARLOS – UFSCar

LICENCIATURA EM MATEMÁTICA

DEPARTAMENTO DE FÍSICA, QUÍMICA E MATEMÁTICA

REGRESSÃO LINEAR E NÃO LINEAR SIMPLES APLICADA À ECONOMIA E FINANÇAS

DELSA KOBATA DOS SANTOS LAHR

Trabalho de Conclusão de Curso apresentado como requisito parcial para a conclusão do Curso Licenciatura em Matemática, sob a Orientação do Prof. Dr. Antônio Luís Venezuela.

Folha de aprovação

Delsa Kobata dos Santos Lahr

"REGRESSÃO LINEAR E NÃO LINEAR SIMPLES APLICADA À ECONOMIA E FINANÇAS"

Trabalho de Conclusão de Curso

Universidade Federal de São Carlos – *Campus* Sorocaba

Sorocaba, 06/07/2017.

Orientador ______________________________
Prof. Dr. Antônio Luís Venezuela

Membro 2 Magda Peixoto ______________________________
Prof.ª Dr.ª Magda da Silva Peixoto

Membro 3 ______________________________
Prof. Dr. Paulo César Oliveira

Ora, àquele que é poderoso para fazer infinitamente mais do que tudo quanto pedimos ou pensamos, conforme o seu poder que opera em nós. (Efésios 3:20).

Agradecimentos

Em primeiro lugar agradeço a Deus por ter me dado forças, ânimo e coragem para chegar até esse momento. Ele foi Aquele que aplainou caminhos, derrubou barreiras, quebrou grandes pedras e permitiu que nesses quatro anos e meio meu caminho fosse uma reta rumo ao alvo, sem me desviar para a direita ou para a esquerda. Somente o querer não é suficiente, não somos donos das situações e condições.

Agradeço ao meu esposo Emerson Amoracil Lahr pela compreensão e paciência, este que acompanhou a minha trajetória, me ajudou em muitas situações tanto nos conteúdos matemáticos como no lar e para que eu desse conta das minhas obrigações abriu mão de muito do meu tempo para com ele.

Agradeço aos meus filhos Karen e Felipe pela compreensão e paciência, pois apesar de me esforçar por estar sempre presente em relação a eles, algum tempo acabou sendo sacrificado.

Agradeço ao professor e orientador tanto na Iniciação Científica como no Trabalho de Conclusão de Curso Antônio Luís Venezuela por toda dedicação, ajuda e paciência que teve comigo.

Agradeço aos professores Graciele Paraguaia Silveira, Paulo Cesar Oliveira e Wladimir Seixas que foram muito especiais em minha trajetória pela UFSCar e que me ajudaram muito.

Agradeço a companheira de sala e amiga muito especial Larissa de Fátima Pastre Tirolla que fez parte da minha trajetória pela UFSCar e me ajudou muito.

Por fim, agradeço a todos os professores e colegas que passaram por mim. É um longo tempo participando de uma comunidade e aprendi muito e uma das aprendizagens é ser um pouco melhor como pessoa. Tive muitas oportunidades de conhecer pessoas no estágio e nas aulas que ministrei e ministro para alunos medalhistas da Olimpíada Brasileira de Matemática, fruto já desse curso que ora finalizo; além do quê vislumbro um futuro diferente, um futuro mais rico.

Sumário

Lista de Tabelas

Lista de Figuras

Resumo

Este trabalho buscou compreender uma técnica de otimização de dados, ou seja, como melhor tratar um conjunto de dados. Para esse objetivo foi utilizado o Método dos Mínimos Quadrados (MMQ), Regressão Linear e não Linear Simples, onde se buscou encontrar o melhor ajuste para um conjunto de dados minimizando a soma dos quadrados das diferenças entre o valor estimado e o valor observado. A fim de testar alguns modelos foram comparados modelos linear e não linear e nesse caso para comparar foi necessário linearizá-las. Foram linearizadas funções do tipo exponencial, potencial e semilogarítmica. No decorrer do trabalho a teoria do MMQ, das Estatísticas de Avaliação R^2, F e t dos parâmetros estimados e Teste de Hipótese são desenvolvidos tanto na forma analítica como na forma computacional, exceto o Teste de Hipótese. Faz-se a Análise de Variância (Anova) e com base nas Estatísticas de Avaliação é possível definir a melhor função estimada e com o Teste de Hipótese é verificado se determinado parâmetro estimado apresenta efeito ou não no resultado final ou fenômeno em análise. Para a aplicação do MMQ de forma analítica trabalhou-se uma tabela oriunda da literatura e computacionalmente com dados reais do SEADE (Fundação Sistema Estadual de Análise de Dados). Para a realização da Análise de Variância e Teste de Hipótese trabalhou-se um modelo da literatura e outro com dados do Instituto Brasileiro de Geografia (IBGE). Pelos resultados observados conclui-se que as Estatísticas de Avaliação em conjunto com o Teste de Hipótese permite encontrar a melhor função estimada para um conjunto de dados.

Palavras-chave: *Método dos Mínimos Quadrados, Regressão Linear, Regressão não Linear Simples, Estatísticas de Avaliação.*

ABSTRACT

This study sought to understand a data optimization technique, namely, how to best treat a DataSet. For this purpose we used the Method of least squares (MMQ), Linear and non Linear regression, Simple, where he sought to find the best fit for a set of data by minimizing the sum of squares of the differences between the estimated value and the observed value. In order to test some models were linear and non linear models compared in this case to compare it was necessary to linearizá them. Were linearized exponential functions, potential and semilogarithmic scale. In the course of work the theory of MMQ, evaluation statistics R^2, F and t of estimated parameters and hypothesis testing are developed both in analytical form as in computational form, except the hypothesis testing. The analysis of variance (Anova) and based on the statistics of evaluation is possible to define the best estimated function and with the hypothesis Test is checked if the given parameter estimated presents the final result or not effect or phenomenon under analysis. For the application of the MMQ analytical form worked out a table from the literature and computationally with actual data SEADE (Foundation State Data Analysis System). For variance analysis and hypothesis testing has been a model of literature and another with data from the Brazilian Institute of geography (IBGE). For the results observed it is concluded that the evaluation Statistics in conjunction with the hypothesis testing allows you to find the best estimated function to a set of data.

Keywords: *Least squares, Linear Regression, nonlinear regression, Simple evaluation Statistics.*

1. Introdução e Objetivos

O tema deste Trabalho de Conclusão de Curso situa-se na área da Matemática Aplicada com a utilização de Regressão Linear e não Linear Simples. Simples porque se estudou o caso em que se têm duas variáveis envolvidas, sendo uma dependente e outra independente. Como principal percurso teórico-metodológico foi utilizado o autor MATOS (2000).

Carl Friedrich Gauss demonstrou em 1809 que minimizar a soma dos quadrados dos resíduos é a melhor maneira de determinar um parâmetro desconhecido de uma equação de condições (aquela que contém uma ou mais observações caracterizadas pelas suas variáveis dependente e independente). Esse método foi chamado mais tarde de Mínimos Quadrados por Adrien-Marie Legendre. Em 1810 foi apresentada a generalização a problemas com mais de uma variável independente por Pierre-Simon Laplace.

O Método dos Mínimos Quadrados (MMQ) é o procedimento de estimação dos parâmetros de um modelo de regressão (representação simplificada de um processo do mundo real) por meio da minimização, como mencionado acima, da soma dos quadrados das diferenças entre os valores observados da variável resposta em uma amostra e seus valores preditos pelo modelo. Possui aplicações em áreas como Biologia, Engenharia, Estatística, Física, Matemática, entre outras, principalmente aquelas que objetivam relacionar uma variável dependente (Y) em função de variáveis explicativas (X_1,..., X_k). Observa-se, por exemplo, o estudo que utiliza estimativas de parâmetros via MMQ a saber: espessura da camada limite de concentração, térmica e fluidos (VENEZUELA, 2008).

Neste trabalhou aplicou-se o conceito de Regressão Linear e não-Linear na área da Econometria e Finanças e para isso alguns conceitos e definições foram explicados previamente a fim de contextualização. A Econometria é definida como sendo um método de análise econômico que agrega a Estatística, a Matemática e a Teoria Econômica. Um modelo econométrico (modelo de regressão) é uma representação simplificada de um processo do mundo real. Assim este modelo é descrito por um conjunto de equações comportamentais derivadas do modelo econômico, as quais envolvem variáveis observáveis e um termo aleatório ou errático, que contém todos os fatores que não foram incorporados ao modelo em análise.

Alguns modelos não são da forma linear e para que avaliações adequadas sejam feitas faz-se necessário linearizar essas funções. Há fenômenos que possuem função do tipo exponencial, potencial, logarítmica e outras; e para que os parâmetros estimados sejam encontrados faz-se necessária a linearização e, depois com a função estimada, encontram-se os valores estimados. Na Econometria a Regressão Linear é amplamente utilizada.

Porém, mesmo os parâmetros estimados devem passar por determinadas verificações e testes para avaliação de sua qualidade. Para isso são calculadas algumas estatísticas de avaliação e nesse trabalho se estuda as principais, dentre elas estatísticas F, t e R^2. Para o cálculo dessas estatísticas faz-se necessário decompor a variação de Y (variável dependente) utilizando a análise de variância (Anova). Essa decomposição permite o cálculo das estatísticas de avaliação referidas anteriormente.

Considerando um intervalo de confiança comumente utilizado para níveis de probabilidade de 95% realiza-se o Teste de Hipótese onde se verifica a presença de efeito ou não do parâmetro estimado no fenômeno estudado. Para isso, se comparam valores encontrados nas estatísticas F e t desses parâmetros com valores disponíveis em tabelas estatísticas.

O desenvolvimento teórico do Método dos Mínimos Quadrados, da Análise de Variância, das Estatísticas de Avaliação e Teste de Hipóteses será apresentado no decorrer deste trabalho.

Para a estimação dos parâmetros via MMQ será apresentada uma solução analítica e outra utilizando o programa R. Na primeira são utilizados dados provindos da literatura e na segunda dados reais do site SEADE (Fundação Sistema Estadual de Análise de Dados).

Para a avaliação dos modelos estimados são utilizados dois exemplos, o primeiro com dados oriundos da literatura e o segundo com dados do Instituto Brasileiro de Geografia (IBGE). Para os modelos estimados foram utilizadas as funções linear, exponencial, potencial e logarítmica com o objetivo de verificar a que melhor explica o fenômeno estudado. Para essa verificação são utilizadas as estatísticas de avaliação e o teste de hipótese. O objetivo dessa formulação é prever uma forma representativa passível de teste empírico, por meio de estimação, teste e checagem do diagnostico produzido.

Este Trabalho de Conclusão de Curso possui cinco capítulos sendo que a Introdução é o Capítulo 1.

No Capítulo 2 é definido Econometria, o conceito e classificação de modelos econométricos e a especificação de modelos. Na especificação de modelos são apresentadas formas funcionais linearizáveis.

No Capítulo 3 é apresentado o Método dos Mínimos Quadrados como um dos métodos de estimação da Regressão Linear Simples, o coeficiente de correlação de Pearson

r e o coeficiente de determinação R^2. No primeiro coeficiente se verifica o quão bem duas variáveis se correlacionam e no segundo verifica-se o quão bem a reta de regressão se ajusta aos dados da amostra. Neste capítulo, utilizando-se o MMQ calculam-se os parâmetros dos modelos econômicos de forma analítica e computacionalmente.

No Capítulo 4 é feita a avalição dos modelos estimados. É apresentado o quadro de Análise de Variância (ANOVA) que serve de base para o cálculo das estatísticas de avaliação F e t e o Teste de Hipótese dentro de um determinado intervalo de confiança.

No Capítulo 5 é feita a determinação analítica do quadro de Análise de Variância e o Teste de Hipótese, utilizando-se das estatísticas de avaliação calculadas com dados da literatura e a determinação via planilha eletrônica dos mesmos itens com dados do IBGE.

Para a implementação foram utilizados os programas de domínio público SciDAVis e Programa R (R Project) e o Excel. Os gráficos foram plotados em SciDAVis, os parâmetros estimados e o R^2 no Programa R, as estatísticas F e t no Excel e o formato do quadro de exibição da Anova similar ao que o Excel apresenta.

Objetivo Geral:

O objetivo geral deste trabalho é aplicar Regressão Linear e não Linear Simples na área de Economia e Finanças.

Objetivos Específicos:

- Desenvolver teoricamente o Método dos Mínimos Quadrados;
- Aplicar o Método dos Mínimos Quadrados analiticamente e computacionalmente;
- Desenvolver teoricamente as estatísticas de avaliação;
- Linearizar as formas funcionais exponencial, potencial e logarítmica;
- Aplicar a Análise de Variância analiticamente e computacionalmente;
- Aplicar Teste de Hipótese nas estatísticas F e t obtidos na Análise de Variância.

2. Econometria

Segundo MATOS (2000), Econometria é o ramo da Economia que trata da mensuração de relações econômicas, isto é, relação entre variáveis de natureza econômica. É uma combinação de teoria com Matemática e Estatística, com o objetivo de dar conteúdo empírico às formulações teóricas da Economia.

De modo específico, são propósitos da Econometria:

- a mensuração de variáveis e agregados econômicos;
- a estimação de parâmetros de relações estabelecidas pela teoria econômica ou outro conhecimento *a priori* ;
- a formulação e teste de hipóteses sobre o comportamento da realidade;
- a previsão de valores de variáveis econômicas.

2.1 Conceito e Classificação de Modelos Econométricos

A palavra *modelo*, de acordo com MATOS (2000) pode ser entendida como uma representação da realidade, estruturada de forma tal que se permita compreender o funcionamento total ou parcial dessa realidade. São estabelecidos um conjunto de hipóteses sobre o comportamento de um fenômeno, com base numa teoria já existente ou a partir de novas proposições teóricas e busca-se entender o fenômeno estudado.

Os modelos econômicos são divididos em teóricos ou econométricos. Nos modelos teóricos não há necessidade da forma matemática ser apresentada; já nos modelos econométricos é necessário que a especificação efetiva da forma matemática compareça para aplicação empírica.

O modelo econométrico também chamado de probabilístico não admite relações exatas, tendo em vista a impossibilidade de incluir todas as variáveis que determinam o comportamento da realidade ou fenômeno além da existência dos erros de medidas das variáveis.

Destacam-se três aspectos dos modelos econométricos: estrutura, classificação quanto às características dos fenômenos e qualidades desejáveis.

Quanto à estrutura envolve quatro elementos básicos: variáveis, relações ou equações, parâmetros ou coeficientes, termo aleatório ou perturbações aleatórias.

Quanto à classificação pode ser subdividido em relação à forma funcional: linear ou não linear; ao número de equações (uniequacionais ou multiequacionais), à associação das variáveis com o tempo (estáticos ou dinâmicos) e à finalidade (modelos de decisão ou de previsão).

Quanto às qualidades desejáveis um modelo econométrico deve possuir plausibilidade teórica, capacidade explanatória, exatidão das estimativas dos parâmetros, capacidade de previsão e simplicidade.

Seguem as etapas para a pesquisa em Economia:

1. Formulam-se hipóteses sobre o comportamento da realidade, as quais são oriundas de uma teoria econômica e/ou observação direta da realidade. Essas hipóteses são reunidas num modelo matemático para sua operacionalização.
2. Coletam-se os dados e estimam-se os parâmetros com a utilização de um método apropriado.
3. Compreende-se a avaliação mediante a utilização de critérios derivados da teoria ou outro raciocínio *a priori*, além de outros de natureza estatística ou econométrica.

2.2 Especificação de Modelos

Segundo MATOS (2000) a especificação de um modelo envolve:

- determinação das variáveis dependentes e explicativas;
- a expectativa *a priori* dos sinais e da magnitude dos parâmetros;
- a forma funcional: linear ou não linear;

2.2.1 Formas funcionais linearizáveis

Segundo MATOS (2000) uma equação do tipo linear significa que as relações entre a variável dependente e a explicativa são lineares; ou seja, qualquer variação (aumento ou redução) registrada na variável explicativa gera uma variação (aumento ou redução) na variável dependente na mesma proporção. Todavia, isso nem sempre acontece. Equações podem ser não lineares ou podem assumir uma combinação de duas ou mais formas. A exigência que se faz é que as relações não lineares sejam linearizáveis por transformação.

Segue um sumário das formas funcionais linearizáveis mais comuns com as respectivas equações originais e transformadas:

Tabela 1 - Formas Funcionais Linearizáveis mais comuns

Tipo da função	**Forma original**	**Forma linearizada por transformação**	**Restrições das variáveis na forma transformada**
Linear	$Y = a + b X$	Nenhuma	Nenhuma
Exponencial	$Y = a . b^x$	$\text{Ln } Y = \text{Ln } a + (\text{Ln } b).X$	$Y > 0$
Potencial	$Y = a . X^b$	$\text{Ln } Y = \text{Ln } a + b. \text{Ln } X$	$Y > 0$ e $X > 0$
Semilogaritmica II	$e^y = a . X^b$	$Y = \text{Ln } a + b. \text{Ln } X$	$X > 0$
Recíproca I	$Y = a + b . \frac{1}{X}$	Usa-se 1/X em vez de X	$X \neq 0$
Recíproca II	$Y = \frac{1}{(a+bX)}$	$1/Y = a + b X$	$Y \neq 0$
Quadrática	$Y = a + bX + cX^2$	Usa-se X^2 além de X	Nenhuma

Fonte: (MATOS, 2000)

2.2.2 Critérios de escolha da forma funcional

De acordo com MATOS (2000), a teoria econômica, em geral, informa muito pouco sobre a forma funcional mais adequada a ser usada na especificação de um modelo econométrico. Cada pesquisador decide pela escolha da forma especificativa dos modelos que formula. Entretanto, existem alguns critérios gerais que norteiam a escolha da forma funcional, a saber:

- Simplicidade – entre uma forma funcional simples e uma complexa, tende-se a escolher a primeira se ambas explicam o fenômeno igualmente bem.
- Indicação da teoria econômica – como o objetivo de um modelo econométrico é dar conteúdo empírico às formulações teóricas, o uso de várias formas funcionais e a escolha da que apresenta resultados mais satisfatórios, mas sem uma justificativa teórica poderá resultar numa mensuração desprovida de significado econômico.
- Poder preditivo – um modelo econométrico deve ser útil para previsões. Isso significa que a forma funcional deve ajustar-se bem aos dados.

3. Regressão Linear Simples

De acordo com MATOS (2000) e GUJARATI (2000), uma vez especificado o modelo, parte-se para a estimação dos parâmetros. O modelo linear simples é aquele que contém apenas uma variável explicativa. A equação é denominada modelo de regressão linear simples. Sua equação básica é a seguinte:

$$Y_i = a + bX_i + u_i \quad (i = 1, 2, ..., n)$$

onde Y_i é a variável dependente, X_i é a variável explicativa, u_i é o termo aleatório, a e b são os parâmetros a serem estimados, n indica o tamanho da amostra e o índice i refere-se à unidade de observação dos valores das variáveis.

Os parâmetros da equação poderão ser estimados a partir de valores amostrais das variáveis Y_i e X_i.

3.1 Método de estimação: Método dos Mínimos Quadrados

A literatura sugere diversos métodos para estimar os parâmetros de um modelo econométrico. Os principais são o método dos mínimos quadrados e o método da máxima verossimilhança. Neste trabalho será tratado o método dos mínimos quadrados.

A equação estimada correspondente ao modelo teórico pode ser indicada da seguinte forma:

$$Y_i = \hat{a}_0 + \hat{a}_1 X_i + e_i$$

sendo $\hat{a}_0$ e $\hat{a}_1$ as estimativas e e_i são os erros, ou resíduos estimados.

O objetivo da estimação pelo Método dos Mínimos Quadrados é obter os parâmetros $\hat{a}_0$ e $\hat{a}_1$, a partir de uma amostra dos valores de Y_i e X_i, de modo que os erros ou resíduos sejam mínimos.

3.1.1 Método dos Mínimos Quadrados (Caso Discreto)

Dados os pontos (X_i, Y_i) $i = 0, 1, 2, ..., n$, pretende-se encontrar uma função da forma:

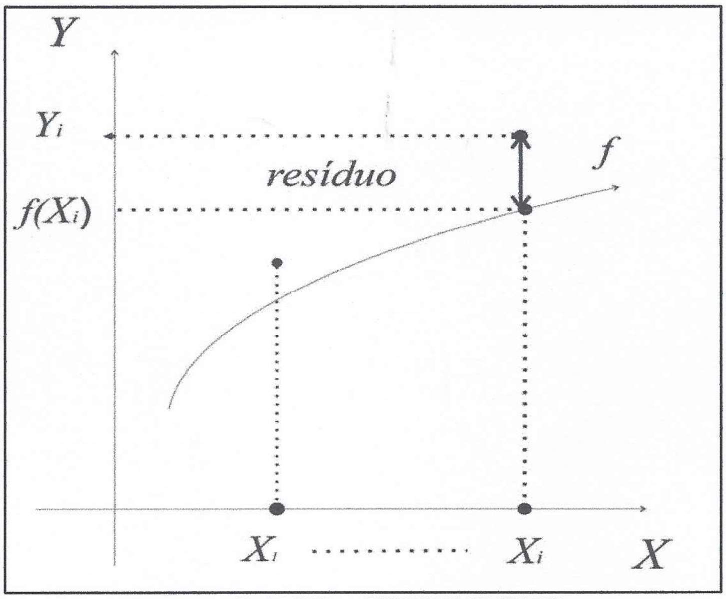
$$f(X) = \sum_{j=0}^{m} a_j f_j(X), \ n \geq m+1 \tag{1}$$

sendo f_j funções continuas e a_j coeficientes reais, tal que a função f, escolhida *a priori*, minimize o funcional:

$$M(a_0, \dots, a_n) = \sum_{i=0}^{n} \left(Y_i - f(X_i)\right)^2 \tag{2}$$

A distância entre Y_i e $f(X_i)$ é chamada de *resíduo* ou *erro* como se vê na Figura 1, um ajuste padrão para o caso discreto.

Figura 1 - Ponto (X_i, Y_i) e função de ajuste com respectivo resíduo



Fonte: SPIEGEL e STEPHENS (2009)

Assim, têm-se pontos conhecidos, (X_i, Y_i), $i = 0,1,2,\dots,n$, e pretende-se aproximá-los para uma função qualquer, por exemplo: linear, exponencial, periódica, etc.

Para que o funcional M seja mínimo, faz-se:

$$\frac{\partial M}{\partial a_j} = 0, \quad j = 0,1,2,\dots,m \tag{3}$$

Na Equação (2) aplica-se a derivada parcial com relação a a_j, $j = 0,1,2,\dots,m$, e obtém-se:

$$\frac{\partial M}{\partial a_j} = \frac{\partial}{\partial a_j} \sum_{i=0}^{n} [Y_i - f(X_i)]^2$$

Aplicando as regras de derivação, obtém-se:

$$\frac{\partial M}{\partial a_j} = -2\sum_{i=0}^{n} [Y_i - f(X_i)] \frac{\partial}{\partial a_j} f(X_i)$$

Substituindo a Equação (1) na expressão acima, tem-se:

$$\frac{\partial M}{\partial a_j} = -2\sum_{i=0}^{n}[Y_i - f(X_i)] \cdot \frac{\partial}{\partial a_j}\sum_{k=0}^{m} a_k f_k(X_i).$$

Assim

$$\frac{\partial M}{\partial a_j} = -2\sum_{i=0}^{n}[Y_i - f(X_i)]\sum_{k=0}^{m} f_k(X_i)\frac{\partial a_k}{\partial a_j} \tag{4}$$

Como $\delta_{kj} = \frac{\partial a_k}{\partial a_j}$, sendo δ_{kj} o delta de Kronecker, o qual é definido por:

$$\delta_{kj} = \begin{cases} 0, \text{ se } k \neq j \\ 1, \text{ se } k = j \end{cases} \tag{5}$$

Com isto, a partir da Equação (4), obtém-se:

$$\frac{\partial M}{\partial a_j} = -2\sum_{i=0}^{n}[Y_i - f(X_i)]f_j(X_i)$$

Considerando a Equação (3) e substituindo a Equação (1) na expressão acima, tem-se

$$0 = -2\sum_{i=0}^{n}[Y_i - \sum_{k=0}^{m} a_k f_k(X_i)]f_j(X_i), \text{ com } j = 0,1,2,...,n. \tag{6}$$

Aplicando lei do cancelamento do produto, o ponto de mínimo do funcional *M*, em relação a a_j, é descrito como:

$$0 = -\sum_{i=0}^{n} Y_i f_j(X_i) + \sum_{i=0}^{n}\sum_{k=0}^{m} a_k f_k(X_i) f_j(X_i)$$

Daí, tem-se:

$$\sum_{i=0}^{n}\sum_{k=0}^{m} a_k f_k(X_i) f_j(X_i) = \sum_{i=0}^{n} Y_i f_j(X_i)$$

Para cada $j = 0,1,2,...,m$, obtém-se o sistema.

$$\begin{cases} \sum_{i=0}^{n}\sum_{k=0}^{m} a_k f_k(X_i) f_0(X_i) = \sum_{i=0}^{n} Y_i f_0(X_i) \\ \sum_{i=0}^{n}\sum_{k=0}^{m} a_k f_k(X_i) f_1(X_i) = \sum_{i=0}^{n} Y_i f_1(X_i) \\ \vdots \\ \sum_{i=0}^{n}\sum_{k=0}^{m} a_k f_k(X_i) f_m(X_i) = \sum_{i=0}^{n} Y_i f_m(X_i) \end{cases} \tag{7}$$

No sistema de equações (7), toma-se o lado esquerdo da equação sendo $j = s$, $j = 0,1,...,m$, e com isto tem-se o seguinte desenvolvimento:

$$\sum_{i=0}^{n}\sum_{k=0}^{m} a_k f_k(X_i) f_s(X_i) = \sum_{i=0}^{n}(a_0 f_0(X_i) f_s(X_i) + ... + a_m f_m(X_i) f_s(X_i))$$

Obtendo assim o seguinte:

$$\sum_{i=0}^{n}\sum_{k=0}^{m} a_k f_k(X_i) f_s(X_i) = a_0 \sum_{i=0}^{n} f_0(X_i) f_s(X_i) + ... + a_m \sum_{i=0}^{n} f_m(X_i) f_s(X_i) \tag{8}$$

Substituindo a Equação (8) no sistema de equações (7), tem-se:

$$\begin{cases} a_0 \sum_{i=0}^{n} f_0(X_i) f_0(X_i) + ... + a_m \sum_{i=0}^{n} f_m(X_i) f_0(X_i) = \sum_{i=0}^{n} Y_i f_0(X_i) \\ a_0 \sum_{i=0}^{n} f_0(X_i) f_1(X_i) + ... + a_m \sum_{i=0}^{n} f_m(X_i) f_1(X_i) = \sum_{i=0}^{n} Y_i f_1(X_i) \\ \vdots \\ a_0 \sum_{i=0}^{n} f_0(X_i) f_m(X_i) + ... + a_m \sum_{i=0}^{n} f_m(X_i) f_m(X_i) = \sum_{i=0}^{n} Y_i f_m(X_i) \end{cases}$$

Como são dados os pontos (X_i, Y_i), $i = 0,1,2,...,n$, e as funções contínuas, f_j, $j = 0,1,2,...,m$, motivado pelo problema fenômeno estudado, determinam-se os coeficientes reais a_j. O sistema acima pode ser escrito na forma matricial como segue:

$$\underbrace{\begin{bmatrix} \sum_{i=0}^{n} f_0(X_i) f_0(X_i) & ... & \sum_{i=0}^{n} f_m(X_i) f_0(X_i) \\ \sum_{i=0}^{n} f_0(X_i) f_1(X_i) & ... & \sum_{i=0}^{n} f_m(X_i) f_1(X_i) \\ & \vdots & \\ \sum_{i=0}^{n} f_0(X_i) f_m(X_i) & ... & \sum_{i=0}^{n} f_m(X_i) f_m(X_i) \end{bmatrix}}_{=A} . \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}}_{=X} = \underbrace{\begin{bmatrix} \sum_{i=0}^{n} Y_i f_0(X_i) \\ \sum_{i=0}^{n} Y_i f_1(X_i) \\ \vdots \\ \sum_{i=0}^{n} Y_i f_m(X_i) \end{bmatrix}}_{=b}, \tag{9}$$

ou seja, $A.X = b$, sendo A a matriz do sistema, X a matriz coluna das incógnitas e b matriz coluna dos termos independentes.

3.1.2 Coeficiente de correlação de Pearson (r)

Correlação é uma medida de associação bivariada (força) do grau de relacionamento entre duas variáveis quantitativas. O coeficiente de correlação de Pearson, r, é uma medida de associação linear entre duas variáveis quantitativas.

Para o seu cálculo são levados em conta a variância e a covariância. A variância é uma medida de dispersão que indica a regularidade de um conjunto de dados em função da média aritmética e a covariância indica o grau de variação conjunta apresentada por duas variáveis.

Seja ρ o coeficiente de correlação entre X e Y como abaixo:

$$\rho = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}} \tag{10}$$

Dada uma amostra de pares de dados (x_n, y_n), $n = 1, ..., N$, obtém-se o coeficiente de correlação amostral substituindo a covariância e as variâncias em (10) por seus correspondentes amostrais:

$$r = \frac{c\hat{o}v(X,Y)}{\sqrt{v\hat{a}r(X)v\hat{a}r(Y)}} \tag{11}$$

sendo:

$$c\hat{o}v(X,Y) = \sum_{t=1}^{N}(x_t - \bar{x})\ (y_t - \bar{y}) / (N-1) \tag{12}$$

$$v\hat{a}r(X) = \sum_{n=1}^{N}(x_t - \bar{x})^2 / (N-1) \tag{13}$$

sendo $\bar{x}$ e $\bar{y}$ = ponto das médias

Define-se a variância amostral de Y da mesma forma que $v\hat{a}r(X)$. Pode-se, pois, escrever o coeficiente de correlação amostral r conhecido como o *coeficiente de correlação de Pearson*:

$$r = \frac{\left(\sum_{i=1}^{N} \Delta x_i \Delta y_i\right)}{\sqrt{\sum_{1}^{N}(\Delta x_i)^2 \sum_{1}^{N}(\Delta y_i)^2}} = \frac{N\left(\sum_{i=1}^{N} x_i y_i\right) - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)}{\sqrt{\left[N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2\right] \cdot \left[N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right]}} \quad (14)$$

onde tem-se o intervalo $-1 \leq r \leq 1$. Portanto pode-se avaliar a correlação entre os dados de duas variáveis qualitativamente, como se descreve a seguir:

Tabela 2 – Correlação Linear no intervalo [0,1]

$0,0 < r < 0,3$	Correlação Linear Fraca
$0,3 \leq r < 0,6$	Correlação Linear Moderada
$0,6 \leq r < 0,9$	Correlação Linear Forte
$0,9 \leq r < 1,0$	Correção Linear Muito Forte

Fonte: SPIEGEL & STEPHENS (2009)

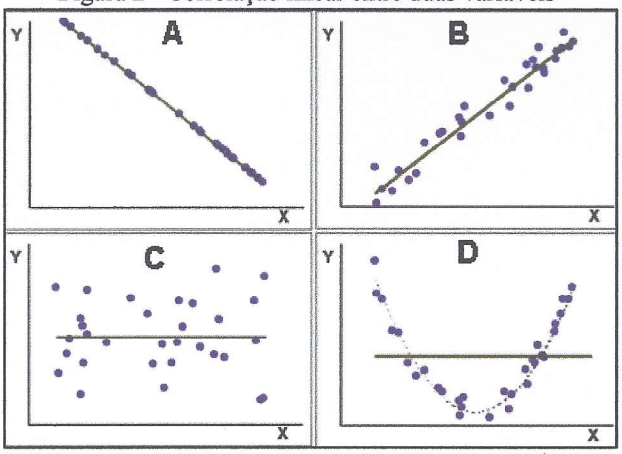
e

Tabela 3 – Correlação Linear no intervalo [-1,0]

$-1,0 < r < -0,9$	Correlação Linear Muito Forte Negativa
$-0,9 \leq r < -0,6$	Correlação Linear Forte Negativa
$-0,6 \leq r < -0,3$	Correlação Linear Moderada Negativa
$-0,3 \leq r < 0,0$	Correlação Fraca Negativa

Fonte: SPIEGEL & STEPHENS (2009)

Figura 2 - Correlação linear entre duas variáveis



Na Figura 2 é possível observar alguns exemplos de correlação entre duas variáveis: correlação linear negativa (A), correlação linear positiva (B). Nas imagens C e D a correlação linear é zero ($r = 0$).

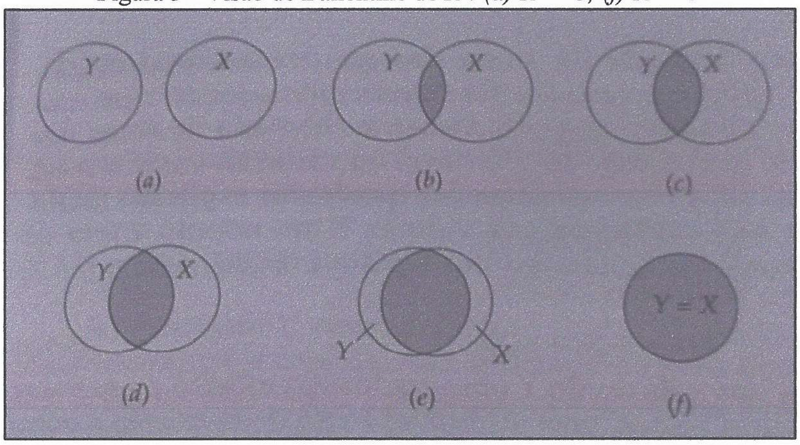
3.1.3 Coeficiente de Determinação (R^2)

Em GUJARATI (2000), o coeficiente de determinação r^2 (regressão simples) ou R^2, sendo (regressão múltipla) é uma medida sintética que diz o quão bem a reta de regressão da amostra se ajusta aos dados.

O coeficiente de determinação R^2 mostra o quão bem a reta de regressão da amostra se ajusta aos dados e r mostra o quanto duas varáveis se correlacionam. Considerando que a reta terá a representação das diversas variáveis explicativas verifica-se o quão bem essas variáveis se correlacionam-se com a variável dependente. Sendo r um valor entre [-1,1] e R^2 ser o coeficiente de correlação de Pearson elevado ao quadrado seu valor será sempre positivo entre [0,1]. Neste trabalho trabalhou-se com uma variável.

Segue uma explicação de R^2 sob o aspecto de um artifício gráfico conhecido como Diagrama de Venn ou *Ballentine* como mostra a Figura 3.

Figura 3 - Visão de Ballentine de R^2: *(a)* $R^2 = 0$; *(f)* $R^2 = 1$



Fonte: GUJARATI (2000)

Na Figura 3, o círculo Y representa a variação na variável dependente Y, o círculo X representa a variação na variável explicativa X. A sobreposição dos dois círculos (área sombreada) indica até que ponto a variação em Y é explicada pela variação X. Quanto maior a sobreposição, maior a variação em Y explicada por X. O R^2 é simplesmente uma medida numérica dessa sobreposição. Na figura, a área sobreposta é crescente da esquerda para a direita, ou seja, uma proporção cada vez maior da variação em Y explicada por X. Em suma, R^2 aumenta. Quando não há sobreposição, R^2 obviamente é zero. Mas quando a sobreposição é completa, R^2 é 100% da variação em Y explicada por X; ou seja, R^2 situa-se entre 0 e 1.

O coeficiente de determinação, R^2, é dado pela seguinte razão:

$$R^2 = \frac{\text{variação explicada}}{\text{variação total}} = \frac{\sum_{i=1}^{n}(\bar{y}-Y_i)^2}{\sum_{i=1}^{n}(\bar{y}-y_i)^2} = \frac{SQE}{SQT}, \qquad (15)$$

onde:

SQR = Soma dos quadrados dos resíduos

SQE = Soma dos quadrados explicada

SQT = Soma dos quadrados total: $SQT = SQR + SQE$

3.2 Aplicação do Método dos Mínimos Quadrados

A aplicação será estabelecida por meio de dados provindos da literatura envolvendo Econometria aplicada às Finanças e de dados reais obtidos em site de instituição oficial de coleta de dados.

Na sequencia inserimos duas aplicações do MMQ. As aplicações 1 e 2 terão seus parâmetros calculados de forma analítica e via planilha eletrônica (computacionalmente).

Na aplicação 1 será utilizado um exemplo disponível em GUJARATI (2000) considerando a função consumo keynesiana[1] linear.

Na aplicação 2 os dados foram obtidos do banco de dados da Fundação Sistema Estadual de Análise de Dados (SEADE).

[1] Keynes declarou: "A lei psicológica fundamental... é que homens [mulheres], como regra e na média, se dispõem a aumentar seu consumo quando sua renda aumenta, mas não tanto quanto o aumento em sua renda".

3.2.1 Aplicação 1: Determinação Analítica e via Planilha Eletrônica dos parâmetros

APLICAÇAO 1: Consumo versus Renda

Determinação Analítica dos parâmetros

Segue tabela com dados hipotéticos sobre consumo familiar semanal C e renda familiar R.

Tabela 4. Consumo familiar semanal e renda familiar, sendo $n = 10$

Consumo (US$) ($C$)	Renda (US$) ($R$)
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

Fonte: GUJARATI (2000)

A partir da Equação (9) considera-se m=1, tem-se a função estimativa:

$$f\left(R\right)=\sum_{i=0}^{1}a_i f_i(R) \qquad (16)$$

Tomando-se f sendo uma função linear, tem-se $f_0\left(R\right)=1$ e $f_1\left(R\right)=R$. Daí:

$$f\left(R\right)=a_0+a_1 R \qquad (17)$$

Logo, a Equação (9) fica escrita como:

$$\begin{bmatrix} \sum_{i=0}^{9} f_0^2(R_i) & \sum_{i=0}^{9} f_0(R_i)f_1(R_i) \\ \sum_{i=0}^{9} f_0(R_i)f_1(R_i) & \sum_{i=0}^{9} f_1^2(R_i) \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{9} C_i f_0(R_i) \\ \sum_{i=0}^{9} C_i f_1(R_i) \end{bmatrix} \qquad (18)$$

Na Tabela 5, têm-se os cálculos necessários para se resolver o sistema acima, nas incógnitas a_0 e a_1, com base na Tabela 4.

Tabela 5 – Cálculos ajuste linear, variável independente R e variável dependente C

i	C	R	$f_0(R_i)$	$f_1(R_i)$	$f_0(R_i)f_1(R_i)$	$f_0^2(R_i)$	$f_1^2(R_i)$	$C_i f_0(R_i)$	$C_i f_1(R_i)$
1	70	80	1	80	80	1	6400	70	5600
2	65	100	1	100	100	1	10000	65	6500
3	90	120	1	120	120	1	14400	90	10800
4	95	140	1	140	140	1	19600	95	13300
5	110	160	1	160	160	1	25600	110	17600
6	115	180	1	180	180	1	32400	115	20700
7	120	200	1	200	200	1	40000	120	24000
8	140	220	1	220	220	1	48400	140	30800
9	155	240	1	240	240	1	57600	155	37200
10	150	260	1	260	260	1	67600	150	39000
	1110	**1700**	**10**	**1700**	**1700**	**10**	**322000**	**1110**	**205500**

Fonte: GUJARATI (2000)

Na Equação (18) substituímos os valores da Tabela 5:

$$\begin{bmatrix} 10,0 & 1700,0 \\ 1700,0 & 322000,0 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1110,0 \\ 205500,0 \end{bmatrix}$$

Resolvendo o sistema acima, obtém-se: a_0 = 24,4545 e a_1 = 0,5091. Daí tem-se a equação: C = 0,5091,0R + 24,4545,0.

Faz-se o cálculo do coeficiente de correlação de Pearson, utilizando a Equação (14) e os dados da Tabela 6, assim:

Tabela 6. Cálculos para o coeficiente de correlação de Pearson

C	R	$f_1^2(R_i)$	C^2	$C_i f_1(R_i)$
70	80	6400	4900	5600
65	100	10000	4225	6500
90	120	14400	8100	10800
95	140	19600	9025	13300
110	160	25600	12100	17600
115	180	32400	13225	20700
120	200	40000	14400	24000
140	220	48400	19600	30800
155	240	57600	24025	37200
150	260	67600	22500	39000
1110	**1700**	**322000**	**132100**	**205500**

Fonte: GUJARATI (2000)

$$r = \frac{10(205500) - 1110(1700)}{\sqrt{[10(132100) - (1110)^2].[10(322000) - (1700)^2]}} = 0{,}9809$$

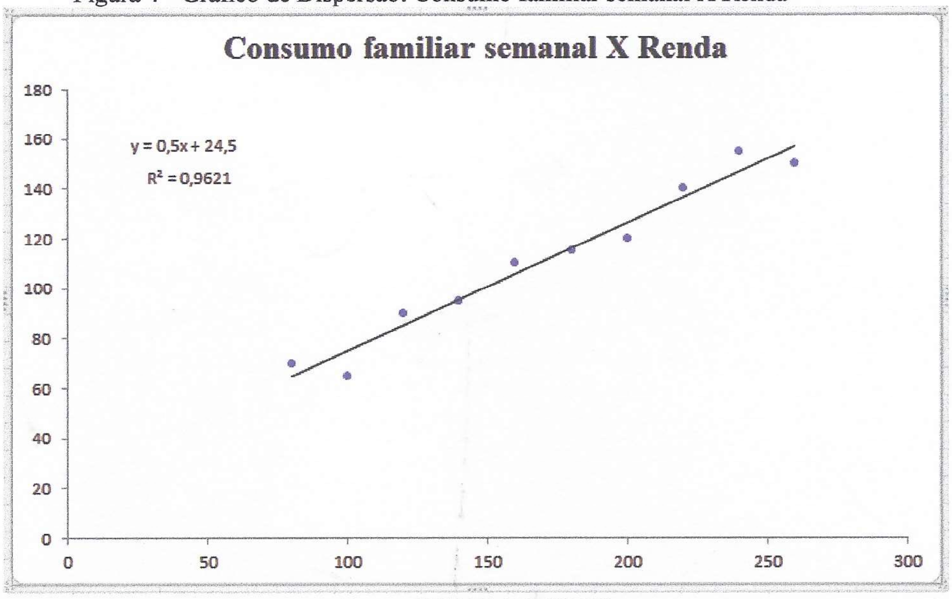
Como r = 0,9809 tem-se que os dados das variáveis R e C tem correlação linear muito forte. Com isto, o modelo linear descreve o comportamento aproximado do consumo C, em função da renda R.

Determinação via planilha eletrônica (computacionalmente) dos parâmetros

A partir da Tabela 4 e utilizando o programa SciDAVis, calculam-se os seguintes parâmetros: $a_0 = 24{,}4545$ e $a_1 = 0{,}5091$, cujo coeficiente de determinação é $r^2 = 0{,}9621$.

Constrói-se o gráfico de dispersão, Figura 4, onde é apresentada a curva de tendência dada pela Equação (15).

Figura 4 - Gráfico de Dispersão: Consumo familiar semanal X Renda



Fonte: GUJARATI (2000)

O coeficiente de correlação de Pearson é r = 0,9809 e isto garante que o modelo tem correlação linear muito forte ($0,9 \leq r < 1,0$) com os dados provindos da Tabela 2. Com isto, o modelo linear descreve o comportamento do consumo familiar semanal em função da renda.

3.2.2 Aplicação 2: Determinação Analítica e via Planilha Eletrônica dos parâmetros

APLICAÇAO 2 – Período versus Consumo de Energia Elétrica Residencial

Determinação Analítica dos parâmetros

Segue determinação analítica dos parâmetros dos modelos econômicos com dados reais:

Tabela 7 - Consumo de Energia Residencial em Sorocaba por ano

Período (P)	Consumo de Energia Elétrica - Residencial (em MWh) (C)
2010	506.993
2011	543.659
2012	586.433
2013	627.615
2014	665.045

Fonte: SEADE

Considerando as Equações (16) e (17), sendo P a variável independente Período e C a variável dependente Consumo a Equação (9) fica escrita como:

$$\begin{bmatrix} \sum_{i=0}^{5} f_0^2(P_i) & \sum_{i=0}^{5} f_0(P_i)f_1(P_i) \\ \sum_{i=0}^{5} f_0(P_i)f_1(P_i) & \sum_{i=0}^{5} f_1^2(P_i) \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{5} C_i f_0(P_i) \\ \sum_{i=0}^{5} C_i f_1(P_i) \end{bmatrix} \quad (19)$$

Na Tabela 8, têm-se os cálculos necessários para se resolver o sistema acima, nas incógnitas a_0 e a_1 com base na Tabela 7.

Tabela 8 - Ajuste linear, variáveis independente P e dependente C, $f_0(P)=1$ $f_1(P)=P$

i	C	P	$f_0(P_i)$	$f_1(P_i)$	$f_0(P_i)f_1(P_i)$	$f_0^2(P_i)$	$f_1^2(P_i)$	$C_i f_0(P_i)$	$C_i f_1(P_i)$
1	506993	2010	1	2010	2010	1	4040100	506993	1019055930
2	543659	2011	1	2011	2011	1	4044121	543659	1093298249
3	586433	2012	1	2012	2012	1	4048144	586433	1179903196
4	627615	2013	1	2013	2013	1	4052169	627615	1263388995
5	665045	2014	1	2014	2014	1	4056196	665045	1339400630
	2929745	**10060**	**5**	**10060**	**10060**	**5**	**20240730**	**2929745**	**5895047000**

Na Equação (19) substituímos os valores da Tabela 8:

$$\begin{bmatrix} 5,0 & 10060,0 \\ 10060,0 & 20240730,0 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 2929745,0 \\ 5895047000,0 \end{bmatrix}$$

Resolvendo o sistema acima, obtém-se: a_0 = -79906123 e a_1 = 40006. Daí tem-se a equação: $C = 40006,0P - 79906123,0$.

Faz-se o cálculo do coeficiente de correlação de Pearson, utilizando a Equação (14) e os dados da Tabela 9, assim:

Tabela 9 - Cálculos para o coeficiente de correlação de Pearson

C	P	$f_1^2(P_i)$	C²	$C_i f_1(P_i)$
506993	2010	4040100	257.041.902.049	1019055930
543659	2011	4044121	295.565.108.281	1093298249
586433	2012	4048144	343.903.663.489	1179903196
627615	2013	4052169	393.900.588.225	1263388995
665045	2014	4056196	442.284.852.025	1339400630
2929745	**10060**	**20240730**	**1732696114069**	**5895047000**

$$r = \frac{5(5895047000) - 2929745(10060)}{\sqrt{[5(1732696114069) - (2929745)^2].[5(20240730) - (10060)^2]}} = 0{,}9997$$

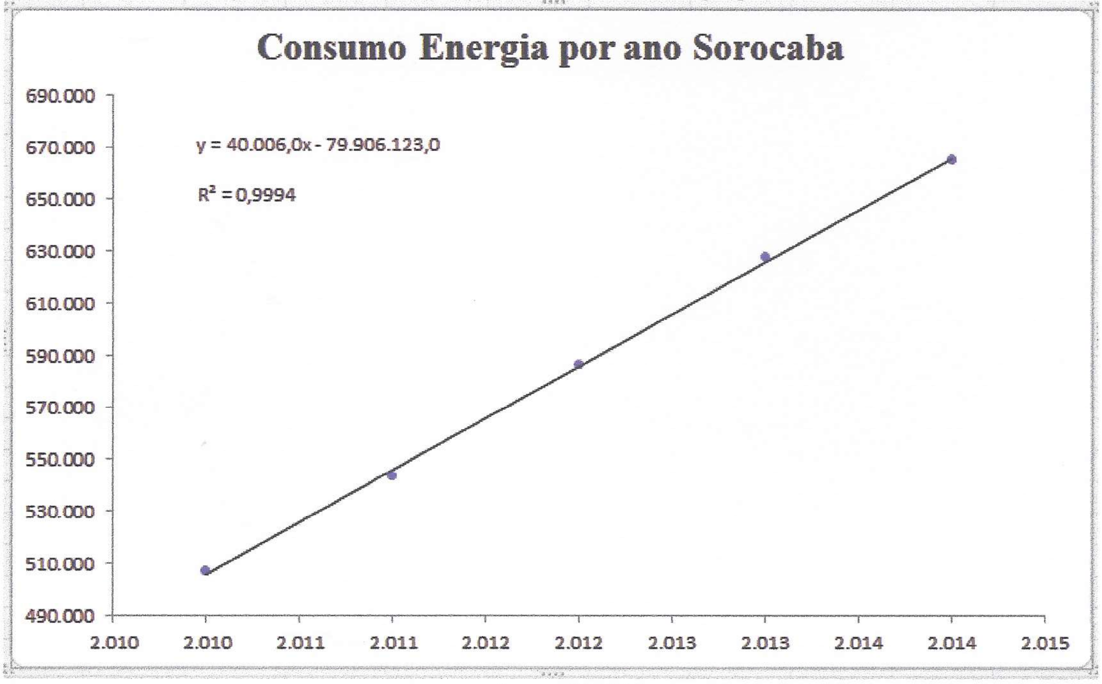
Como r = 0,9997 tem-se que os dados das variáveis P e C tem correlação linear muito forte. Com isto, o modelo linear descreve o comportamento aproximado do consumo C, em função do período P.

Determinação via planilha eletrônica (computacionalmente) dos parâmetros

A partir da Tabela 7 e utilizando o programa SciDAVis, calculam-se os seguintes parâmetros: a_0 = -79906123 e a_1 = 40006, cujo coeficiente de determinação é r^2 = 0,9994.

Constrói-se o gráfico de dispersão, Figura 5, onde é apresentada a curva de tendência dada pela Equação (15).

Figura 5 - Gráfico de Dispersão: Consumo Energia por ano - Sorocaba



Fonte: SEADE

O coeficiente de correlação de Pearson é r = 0,9997 e isto garante que o modelo tem correlação linear muito forte ($0,9 \leq r < 1,0$). Com isto, o modelo linear descreve o comportamento do consumo residencial de energia elétrica em função do período.

Utilizando a equação C = 40006,0P -79906123,0, podemos estimar o consumo de energia em Sorocaba para 2015. Realizando os cálculos tem-se 705967. De acordo com o SEADE houve uma queda de 1,23% no consumo de energia das residências sorocabanas de 2014 para 2015 ficando em 656869.

Em relação ao resultado estimado 705967 e o concretizado 656869 houve uma diferença de 7,475% a maior.

4. Avaliação de Modelos Estimados

De acordo com MATOS (2000) a avaliação da estimativa de um modelo, além da verificação de sua consistência teórica, envolve a mensuração do grau de confiabilidade em termos probabilísticos, sem o que não se conhecerá sua validade para decisões ou previsões.

4.1 Caracterização de uma variável aleatória

De acordo com MATOS (2000), toda variável aleatória é caracterizada por um ou mais parâmetros, que se deseja estimar a partir de uma amostra. Considere que se deseja estimar o parâmetro q.

A fórmula ou procedimento utilizado para transformar um conjunto de dados amostrais em um valor particular chama-se estimador. O valor particular gerado por um estimador é chamado estimativa. Simbolizar-se-á aqui a estimativa do parâmetro q por $\hat{q}$.

Como pode ser obtido a partir de amostras aleatórias repetitivas, a estimativa $\hat{q}$ é, também, aleatória. Ressalte-se que $\hat{q}$ é função de variáveis aleatórias, portanto possui média e variância.

Uma variável aleatória, no plano populacional, possui elementos que a caracterizam, como:

a) Média ou esperança matemática: $E(q)$
b) Variância: $Var(q) = E[\, q - E(q)]^2$
c) Erro-padrão: $EP(q) = \sqrt{Var(q)}$

A comparação dessas características populacionais com as de uma amostra permite a definição de conceitos muito úteis em Econometria a saber, por exemplo, erro amostral $= q - \hat{q}$. Note-se ainda que a esperança matemática de $\hat{q}$, $E(\hat{q})$, é obtida através de repetidas amostras aleatórias.

4.2 Qualidades desejáveis dos estimadores

Segundo MATOS (2000) é desejável que um estimador e a estimativa, $\hat{q}$, que ele gera, tenham certas qualidades, de modo que sejam obtidas as informações mais fidedignas possíveis sobre o valor do verdadeiro parâmetro, q. Isso significa dizer que o desejável é que a distribuição dos valores de $\hat{q}$, obtidos a partir de amostras repetitivas, seja o máximo possível concentrada em torno de q. Para isso, a estimativa $\hat{q}$ deve possuir as qualidades de não-tendenciosidade, eficiência e consistência.

Não-tendenciosidade: se o valor esperado ou médio de $\hat{q}$ for igual a q. Isso significa que, se $\hat{q}$ for calculado um grande número de vezes para amostras distintas, a média de todas essas estimativas será igual a q. Se não for igual, significa que há tendenciosidade.

Eficiência ou variância mínima: um estimador não-tendencioso mas com grande variância, conduzirá a estimativas muito distantes da realidade ou do verdadeiro parâmetro q. Um estimador é eficiente se 1) $\hat{q}$ é uma estimativa não-tendenciosa e 2) $Var(\hat{q}) < Var(\tilde{m})$ onde $\tilde{m}$ é uma estimativa obtida mediante a utilização de outro estimador.

Consistência: referem-se às mudanças que ocorrem na distribuição amostral à medida que o tamanho da amostra cresce. Uma estimativa $\hat{q}$ de q é consistente, se o limite da probabilidade de ocorrência de $\hat{q}$ for igual a q, quando o tamanho da amostra aumentar.

4.3 Média, variância e covariância dos parâmetros estimados

De acordo com MATOS (2000) a avaliação do grau de confiabilidade dos parâmetros estimados de um modelo requer o conhecimento de sua média, variância e covariância. Desenvolvem-se as expressões dos estimadores de mínimos quadrados, que geram as estimativas dos parâmetros $\boldsymbol{a}$ e $\boldsymbol{b}$ do *modelo linear simples* $\boldsymbol{Y} = \boldsymbol{b} + \boldsymbol{a}\boldsymbol{X} + \boldsymbol{e}$, mostrando-se como são derivadas sua média, variância e covariância e como estas são definidas em termos amostrais. Tais fórmulas não serão explicitadas, pois podem ser encontradas em livros de Econometria. [4],[5],[6],[11],[12],[13].

Assim tem-se:

- Média e variância de â
- Média e variância de $\hat{b}$
- Covariância de â e de $\hat{b}$

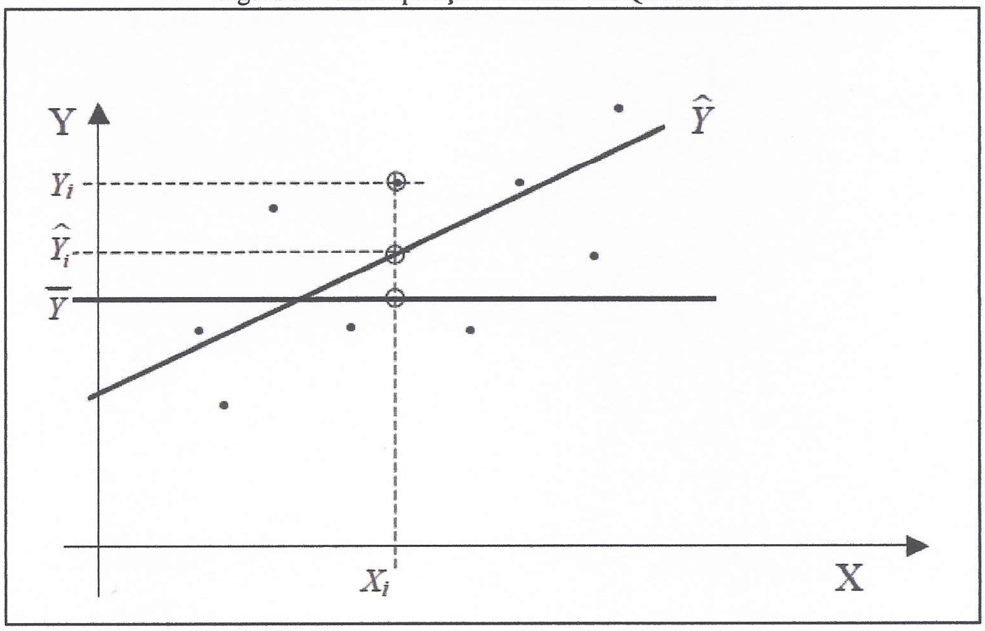
4.4 Decomposição da Soma dos Quadrados

Segundo MATOS (2000) em análise de regressão, é importante que se conheça a decomposição da variação de *Y*, a variável dependente. A variação de *Y (VT)* é definida pelo somatório dos desvios de *Y* em relação a sua média, elevados ao quadrado, isto é:

$$VT = \sum y^2 - \sum (y - \bar{y})^2$$

O interessante agora é decompor *VT* com o objetivo de identificar a parcela explicada pela variável explicativa, *X*. Para isso, considere, inicialmente, um ponto qualquer do espaço *A=(X,Y)* de valores observados. A partir desse ponto, pode-se observar que: $(Y - \bar{Y}) = (\hat{Y} - \bar{Y}) + (Y - \hat{Y})$ ou $y = \hat{y} + e$, com e sendo a estimativa do erro teórico, conforme Figura 4.

Figura 6 – Decomposição da Soma dos Quadrados



Fonte: MATOS (2000)

Elevando-se tal expressão ao quadrado e somando-se seus termos, obtém-se:

$$\sum y^2 = \sum (\hat{y} + e)^2$$

e considerando $\sum \hat{y}e = 0$, obtem-se:

$$\sum y^2 = \sum \hat{y}^2 + \sum e^2$$

e como $\hat{y} = \hat{b}x$ e $\hat{b} = \frac{\sum yx}{\sum x^2}$, tem-se:

$$\sum y^2 = \hat{b}^2 \sum x^2 + \sum e^2 = \hat{b} \sum yx + \sum e^2$$

onde

a) $\sum y^2$ é a variação total de *Y* (*VT*) ou soma de quadrados total (*SQT*).

b) $\hat{b} \sum yx$ é a variação explicada por *X* (*VE*) ou soma de quadrados da regressão (*SQE*).

c) $\sum e^2$ é a variação residual (*VR*) ou soma de quadrados residual (*SQR*).

Portanto, $VT = VE + VR \Leftrightarrow SQT = SQE + SQR$

Tal soma indica que a variação dos valores de *Y* em torno de sua média (*VT*) pode ser decomposta em duas partes: uma que corresponde à de *X* (*VE*) e a outra que expressa a variação residual ou não explicada por *X* (*VR*). O valor de *VR* é atribuído ao fato de que os pontos observados nem sempre pertencem à reta da regressão devido a fatores omitidos ou aleatórios a que estão sujeitas as variáveis econômicas.

4.5 Estatísticas de avaliação

A decomposição da variação de *Y* é normalmente utilizada para a análise de variância, como é mostrado na Tabela 10 abaixo. A análise de variância tem a vantagem de mostrar as diversas relações básicas, úteis no desenvolvimento de estatísticas de avaliação de resultados obtidos com a estimação do modelo.

4.5.1 Análise de Variância Simples (Anova)

Utilizaremos a formatação do Resumo dos Resultados conforme Tabela 10 com dados obtidos pelo programa R (*R Project*) de acordo MATOS (2000). Tal resumo será dado na sequência, onde nos restringiremos apenas às informações úteis à avaliação de modelos lineares e linearizáveis.

Tabela 10: Quadro de Análise de Variância

RESUMO DOS RESULTADOS				$F_c =$ $t_{c,a} =$ $t_{c,b} =$
Estatística de regressão				
$R^2 =$	$\frac{VE}{VT}$			
Observações	N			
ANOVA				
	gl	*SQ*	*MQ*	*F*
Regressão	k	$VE = a\left(\sum YX - \left(\sum Y \sum X\right)\frac{1}{N}\right)$	$\frac{VE}{k}$	$F_{(k,N-k-1)} = \frac{\frac{VE}{k}}{\frac{VR}{N-k-1}}$
Resíduo	$N-k-1$	$VR = VT - VE$	$s^2 = \frac{VR}{(n-K-1)}$	
Total	$N-1$	$VT = \sum Y^2 - \frac{\left(\sum Y\right)^2}{N}$		
	Coeficientes	*Erro padrão*	*Stat t*	
Interseção	a	$S_a = \sqrt{\frac{\sum X^2 s^2}{N \sum x^2}}$	$t_a = \frac{\hat{a}}{S_a}$	
X - Tarifa	b	$S_b = \sqrt{\frac{s^2}{\sum x^2}}$	$t_b = \frac{\hat{b}}{S_b}$	

Fonte: MATOS (2000)

4.5.2 Definição das principais estatísticas de avaliação

Segundo MATOS (2000) o quadro de Resumo dos Resultados nos fornece os elementos que permitem definir as estatísticas de avaliação.

Variância amostral ou residual

$$s^2 = \frac{VR}{(n-K-1)} = \frac{\sum e^2}{(n-K-1)} = \frac{\sum\left(Y-\hat{Y}\right)^2}{(n-K-1)}$$

onde $(n-K-1)$ é o número de graus de liberdade do estimador de mínimos quadrados da variância amostral s^2.

A estimativa da variância residual, s^2, mede o grau de dispersão entre os valores observados e os estimados de *Y*. A raiz quadrada de s^2 é denominada erro-padrão da estimativa, isto é, $s = \left(s^2\right)^{\frac{1}{2}}$.

Estatística *F*

$$F_{(k,N-k-1)} = \frac{\frac{VE}{k}}{\frac{VR}{N-k-1}}$$

A estatística *F* tem por finalidade testar o efeito conjunto das variáveis explicativas sobre a dependente. Significa verificar se, pelo menos, uma das variáveis explicativas do modelo exerce alguma influência sobre a variável dependente. É claro que, no caso do modelo linear simples por existir apenas uma variável explicativa, a função da estatística *F* é a de testar significância do efeito *X* sobre *Y*.

Estatística *t*

A estatística *t* tem por finalidade testar a significância dos parâmetros estimados do modelo, o que equivale ao teste do efeito individual de *X* e do termo constante. Portanto, é definida para cada um dos parâmetros estimados.

1. Estatística *t* para a estimativa do parâmetro *b*

$$t_b = \frac{\hat{b}-b}{S_b}, \text{ ou } \frac{\hat{b}}{S_b} \text{ para } b = 0$$

onde: $S_b = \sqrt{\frac{s^2}{\sum x^2}}$

No caso do modelo linear simples, $t_b = (F)^{1/2}$. Portanto, um teste com a estatística *t* equivale, nesse caso, ao realizado com a estatística *F*.

2. Estatística t para a estimativa do parâmetro a (termo constante)

$$t_a = \frac{\hat{a} - a}{S_a} \text{, ou } \frac{\hat{a}}{S_a} \text{ para } a = 0$$

$$\text{onde: } S_a = \sqrt{\frac{\sum X^2 s^2}{N \sum x^2}}$$

4.6 Teste de Hipóteses e intervalos de confiança

Segundo MATOS (2000) os conceitos de testes de hipóteses e de intervalos de confiança estão intimamente relacionados. Ambos serão desenvolvidos aqui com o objetivo de auxiliar a avaliação de estimativas de modelos econométricos.

4.6.1 Teste de hipóteses

Segundo MATOS (2000) um teste de hipótese é um processo capaz de afirmar, com base em dados amostrais, se uma hipótese sob prova é correta ou não. É uma afirmação ou enunciado que admite se certo efeito está presente ou não. Tem por objetivo auxiliar na decisão acerca de parâmetros populacionais, a partir de informações amostrais.

Tipos de testes - Para avaliar a confiabilidade estatística da estimativa de um modelo, são usados fundamentalmente os testes F e t, que envolvem respectivamente as distribuições F e t Student.

Teste F – O objetivo do teste F é verificar a significância estatística do efeito conjunto das variáveis explicativas de um modelo. Chamaremos de F calculado (F_{calc}), cuja fórmula para tal cálculo foi apresentada anteriormente. Já o F crítico (F_{crit}), pode ser obtido através de tabelas estatísticas (vide Anexo). Assim se tem as hipóteses:

H_0: a = 0 (ausência de efeito)

H_1: a $\neq$ 0 (presença de efeito positivo ou negativo) ou

H_1: a > 0 (presença de efeito positivo) ou

H_1: a < 0 (presença de efeito negativo)

Regra de decisão: Se $F_{calc} \geq F_{crit}$, então Rejeita H_0

Teste *t* – Tem a finalidade de verificar se o efeito de uma variável e a contribuição do termo constante é, ou não, relevante em termos de confiabilidade. Chamaremos de t calculado, respectivamente, para o coeficiente angular "a" e o coeficiente linear "b" (t_a ou t_b), cuja fórmula para tal cálculo foi apresentada anteriormente. Já o t crítico para o coeficiente "a" e "b", respectivamente ($t_{c,a}$ ou $t_{c,b}$), pode ser obtido através de tabelas, ou por intermédio de comando do Excel. Assim se tem as hipóteses:

H_0: a = 0 (ausência de efeito)

H_1: a $\neq$ 0 (presença de efeito positivo ou negativo) ou

H_1: a > 0 (presença de efeito positivo) ou

H_1: a < 0 (presença de efeito negativo)

Regra de decisão: Se $|t_a| \geq |t_{c,a}|$, então Rejeita H_0 e

Se $|t_b| \geq |t_{c,b}|$, então Rejeita H_0

4.6.2 Intervalos de confiança

De acordo com MATOS (2000) pode-se definir intervalos de confiança, dentro dos quais o valor verdadeiro do parâmetro populacional cairá com determinado grau de probabilidade. Comumente trabalha-se com intervalos de confiança cujo nível de probabilidade é de 95%.

Exemplos serão desenvolvidos no programa R.

5. Aplicação da Análise de Variância Simples (Anova)

A aplicação será estabelecida por meio de dados provindos da literatura envolvendo Econometria aplicada às Finanças e de dados reais obtidos no Instituto Brasileiro de Geografia e Estatística (IBGE).

Na sequência será elaborado quadro de análise de variância conforme o Resumo dos Resultados apresentados no item 4.5.1: *VE*, *VT*, *VR* e as estatísticas de avaliação R^2, F, t_b e t_a de forma analítica.

Realizar-se-á o Teste de Hipótese onde o F_crítico e o t_crítico estão disponíveis em tabelas estatísticas, mas pelo fato do programa R (*R Project*) calculá-lo utilizar-se-á dessa facilidade.

Os pontos da amostra serão esboçados num gráfico a fim de se observar a característica de sua curva; em seguida será feita a linearização por transformação de algumas formas funcionais. Neste trabalho optou-se por utilizar os modelos exponencial, potencial e logarítmica. Serão obtidas as estimativas da função da demanda por essas formas funcionais para verificar a que melhor se ajusta aos dados. Os parâmetros das funções citadas acima serão obtidas computacionalmente.

A Anova será gerada computacionalmente e será observado o R^2 e a estatística F que nos dará a informação de qual das quatro funções é a melhor: linear, exponencial, potencial ou recíproca.

Para cada uma das Anovas será realizado o Teste de Hipótese.

Por fim será feita computacionalmente a Análise de Variância em uma amostra de dados reais.

5.1 Determinação Analítica da Anova

Para elaborar o quadro de análise de variância deve-se ter a amostra e estimar a função da demanda. Com a função serão calculados os valores estimados e o erro.

Considere a Tabela 11 disponível em MATOS (2000) onde (*Q*) é o índice de quantidade demandada e (*T*) é a tarifa real média de energia elétrica:

Tabela 11 - Demanda de Energia (Q) em função da Tarifa (T)

Ano	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
Q	69	76	81	90	94	100	103	108	113	115
T	143	134	117	111	109	100	137	122	85	90

Fonte: MATOS (2000)

A estimação da função de demanda consiste em calcular os parâmetros do modelo pelo método dos mínimos quadrados.

A partir da Tabela 11 e utilizando o aplicativo R (*R Project*), calculam-se os seguintes parâmetros: a = 158,805 e b = - 0,557 assim $\hat{Q} = 158{,}805 - 0{,}557T$. Substituindo-se os valores de T na função estimada, obtêm-se as estimativas de Q e, a partir destas, calculam-se os resíduos ou erros.

Tabela 12 - Estimativas de Q e cálculo do erro

Ano	Q_i	$\hat{Q}_i$	$e_i = Q_i - \hat{Q}_i$
1981	69	79,20	-10,20
1982	76	84,21	-8,21
1983	81	93,68	-12,68
1984	90	97,02	-7,02
1985	94	98,13	-4,13
1986	100	103,14	-3,14
1987	103	82,54	20,46
1988	108	90,89	17,11
1989	113	111,49	1,51
1990	115	108,71	6,29
Total	949	949,00	0,00

Fonte: MATOS (2000)

Calculando-se as informações requeridas, obtêm-se os elementos para a elaboração do quadro de análise de variância que servirá de base para o cálculo das estatísticas de avaliação. Seguem os cálculos:

a)

$$VE = a\left(\sum YX - \left(\sum Y \sum X\right)\frac{1}{N}\right)$$

$$VE = -0,557\left(107006 - (949.1148)\frac{1}{10}\right)$$

$$VE = (-0,557).(-1939,2) = 1079,55$$

b)

$$VT = \sum Y^2 - \frac{\left(\sum Y\right)^2}{N}$$

$$VT = 92301 - (949)^2 / 10 = 2240,90$$

c)

$$VR = VT - VE$$

$$VR = 2240,90 - 1079,55 = 1161,35$$

d)

$$R^2 = \frac{VE}{VT} = \frac{1079,55}{2240,90} = 0,482$$

e)

$$F = F_{(k,N-k-1)} = \frac{\frac{VE}{k}}{\frac{VR}{N-k-1}} \qquad F_{(1,8)} = \frac{\frac{1079,55}{1}}{\frac{1161,35}{(10-1-1)}} = 7,437$$

f)

t_b = parâmetro

$$t_b = \frac{\hat{b}}{S_b} \text{ onde } S_b = \sqrt{\frac{s^2}{\sum x^2}}$$

onde $s^2 = \frac{VR}{(n-K-1)}$ $\qquad s^2 = \frac{VR}{(n-K-1)} = \frac{1161,35}{8} = 145,169$

$$\sum x^2 = \sum X^2 - \left(\sum X\right)^2 / N = 135274 - (1148)^2 / 10 = 3483,60$$

$$S_b = \sqrt{\frac{s^2}{\sum x^2}} = \sqrt{\frac{145,169}{3483,60}} = 0,204$$

$$t_b = \frac{\hat{b}}{S_b} = \frac{-0,557}{0,204} = -2,73$$

g)

t_a = termo constante

$$t_a = \frac{\hat{a}}{S_a} \text{ onde } S_a = \sqrt{\frac{\sum X^2 s^2}{N\sum x^2}} \text{ onde } s^2 = \frac{VR}{(n-K-1)}$$

$$S_a = \sqrt{\frac{\sum X^2 s^2}{N\sum x^2}} = \sqrt{\frac{(135274.145,169)}{(10.3483,60)}} = 23,743$$

$$t_a = \frac{\hat{a}}{S_a} = \frac{158,805}{23,743} = 6,69$$

Uma vez obtido os valores das estatísticas de avaliação, restam, por fim, a realização dos testes de hipóteses e a análise dos resultados.

Segue resumo dos resultados obtidos:

$\hat{Q} = 158{,}805 - 0{,}557T$ $\quad R^2 = 0{,}482$

$t_a = 6{,}69$ $\quad t_b = -2{,}73$ $\quad F = 7{,}437$ e $n = 10$

Teste de Hipóteses: Utilizando-se das estatísticas F e t calculadas, realizam-se os Testes de Hipótese.

a) Teste F

Para 1 e 8 graus de liberdade no numerador e denominador, respectivamente, e nível de significância de 5% o valor crítico $F_{crit} = 5{,}32$.

Regra de decisão: Se $F_{Calc} \geq F_{Crit}$, então Rejeita H_0

Regra de decisão: Se $7{,}437 \geq 5{,}32$, então Rejeita H_0 (ausência de efeito), logo presença de efeito.

b) Teste t_a

Para 8 graus de liberdade e nível de significância de 5%, o valor crítico é $t_{c,a}$ = 2,306.

Regra de decisão: Se $|t_a| \geq |t_{C,a}|$, então Rejeita H_0

Regra de decisão: Se $|6{,}69| \geq |2{,}306|$, então Rejeita H_0 (ausência de efeito), logo presença de efeito.

c) Teste t_b

Para 8 graus de liberdade e nível de significância de 5%, o valor crítico é $t_{c,b}$ = 2,306.

Regra de decisão: Se $|t_b| \geq |t_{C,b}|$, então Rejeita H_0

Regra de decisão: Se $|-2{,}73| \geq |2{,}306|$, então Rejeita H_0 (ausência de efeito), logo presença de efeito negativo, de acordo com a teoria da demanda.

Os valores críticos para as estatísticas F e t são encontrados em tabelas estatísticas. Nesse trabalho foram utilizadas as tabelas disponíveis em MATOS (2000).

Análise de Resultados: Considerando somente os critérios teóricos e estatísticos tendo em vista que as estatísticas de verificação de validade das qualidades desejáveis dos estimadores não são tratadas neste trabalho, tem-se:

$\hat{Q} = 158{,}805 - 0{,}557T$ $\qquad R^2 = 0{,}482$

$t_a = 6{,}69 \qquad t_b = -2{,}73 \qquad F = 7{,}437$ e $n = 10$

De acordo com MATOS (2000), apesar de a amostra ser pequena (n = 10), os resultados obtidos são satisfatórios. A variável explicativa tarifa (T) apresenta coeficiente estimado com o sinal teoricamente correto e estatisticamente significativo no nível de 5% de probabilidade de erro, embora tal coeficiente possa ser não-tendencioso em face de omissão de variáveis explicativas importantes. O termo constante é igualmente significativo no nível de 5%.

O teste realizado com estatística F confirma, também, a rejeição da hipótese de efeito nulo da variável preço, o que era de se esperar, dada a equivalência dos testes F e t.

O coeficiente de determinação de 0,482 indica que a variável tarifa real foi responsável por 48,2% da variação da quantidade demandada, Q.

5.2 Determinação via planilha eletrônica (computacionalmente) da Anova

Considere a Tabela 11, deseja-se estimar a equação da demanda obtidas pelas formas funcionais linear, exponencial, potencial e semilogarítmica II e escolher a forma funcional que melhor se ajusta aos dados.

No caso de modelos com apenas uma variável explicativa, a forma funcional mais adequada pode ser escolhida em função da magnitude do coeficiente de determinação (R^2) e da estatística *F*. Quanto mais elevado o valor dessas estatísticas mais forte será a influência da variável explicativa sobre a dependente, ou seja, os dados ajustam-se, mais adequadamente, a forma matemática especificada.

1. Forma Linear: gráfico e quadro de análise (Anova)

Figura 7 - Forma Linear

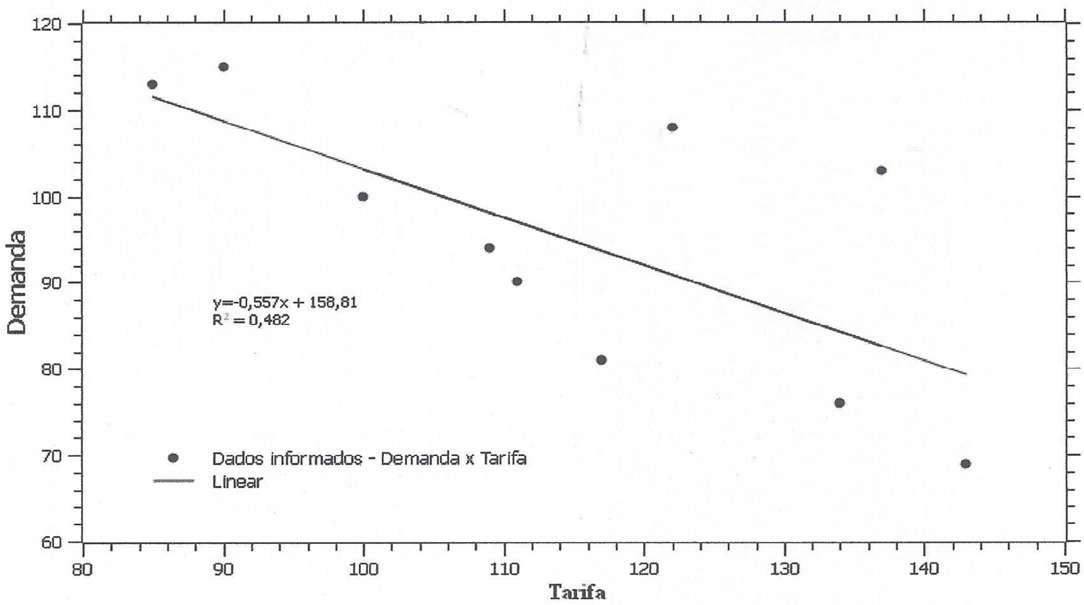


Tabela 13 – Resumo dos Resultados - função linear

			$F_c = 5{,}318$	
Estatística de regressão			$t_{c,a} = 2{,}306$	
$R^2 =$	0,482		$t_{c,b} = 2{,}306$	
Observações	10			
ANOVA				
	gl	*SQ*	*MQ*	*F*
Regressão	1	1079,486	1079,486	7,436
Resíduo	8	1161,414	145,177	
Total	9	2240,9		
	Coeficientes	*Erro padrão*	*Stat t*	
Interseção	158,805	23,743	6,688	
X – Tarifa	-0,557	0,204	-2,727	

2. Forma Exponencial: gráfico e quadro de análise (Anova)

Figura 8 - Forma Exponencial

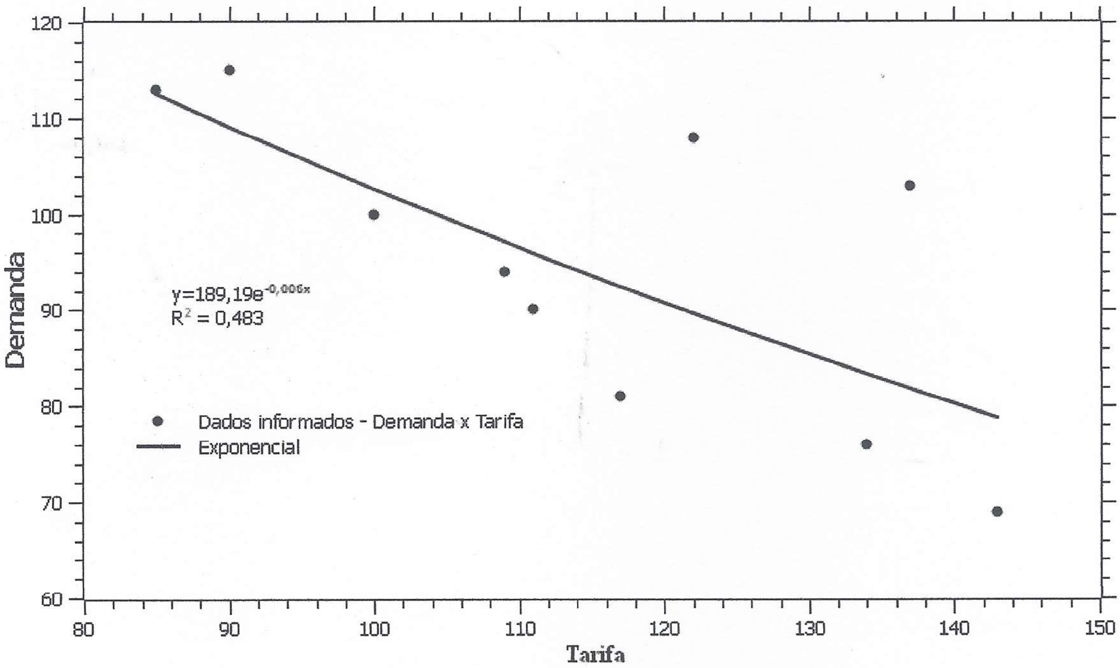


Tabela 14 – Resumo dos Resultados - função exponencial

			$F_c = 5{,}318$	
Estatística de regressão			$t_{a,c} = 2{,}306$	
$R^2 =$	0,483		$t_{b,c} = 2{,}306$	
Observações	10			
ANOVA				
	gl	*SQ*	*MQ*	*F*
Regressão	1	0,131	0,131	7,465
Resíduo	8	0,140	0,017	
Total	9	0,271		
	Coeficientes	*Erro padrão*	*Stat t*	
Interseção	5,243	0,261	20,109	
X - Tarifa	-0,006	0,002	-2,732	

3. Forma Potencial: gráfico e quadro de análise (Anova)

Figura 9 - Forma Potencial

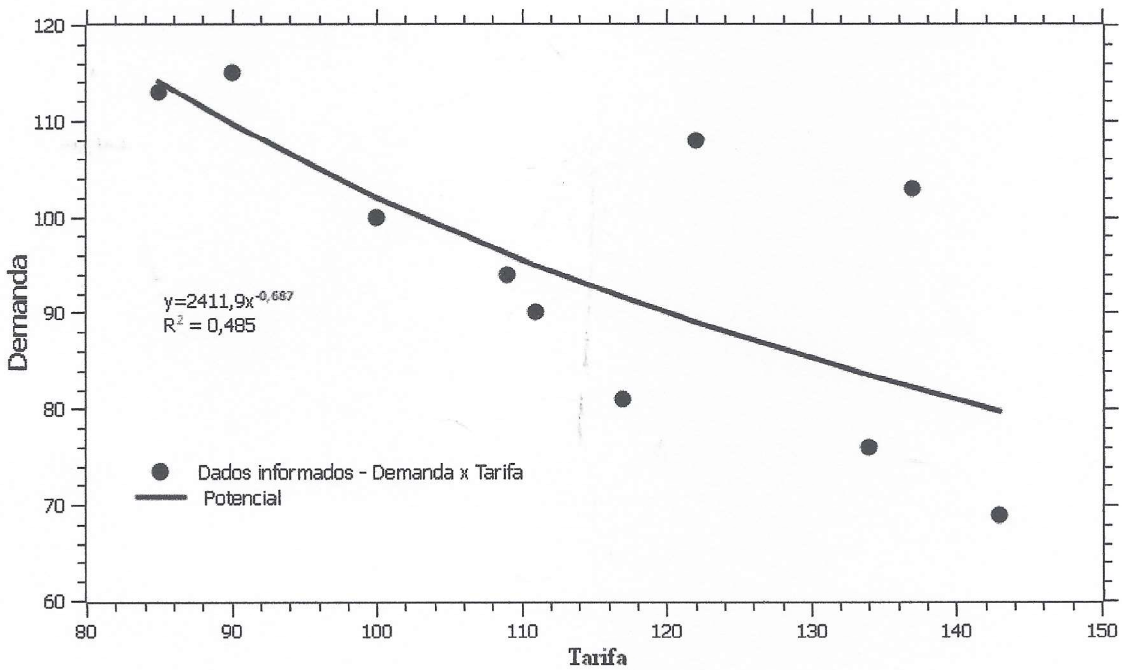


Tabela 15 – Resumo dos Resultados - função potencial

			$F_c = 5{,}318$	
Estatística de regressão			$t_{c,a} = 2{,}306$	
$R^2 =$	0,484		$t_{c,b} = 2{,}306$	
Observações	10			
ANOVA				
	gl	*SQ*	*MQ*	*F*
Regressão	1	0,131	0,131	7,520
Resíduo	8	0,139	0,0174	
Total	9	0,270706864		
	Coeficientes	*Erro padrão*	*Stat t*	
Interseção	7,788	1,185	6,571	
Ln X	0,687	0,250	-2,742	

4. Forma Semilogarítmica II: gráfico e quadro de análise (Anova)

Figura 10 - Forma Semilogarítmica II

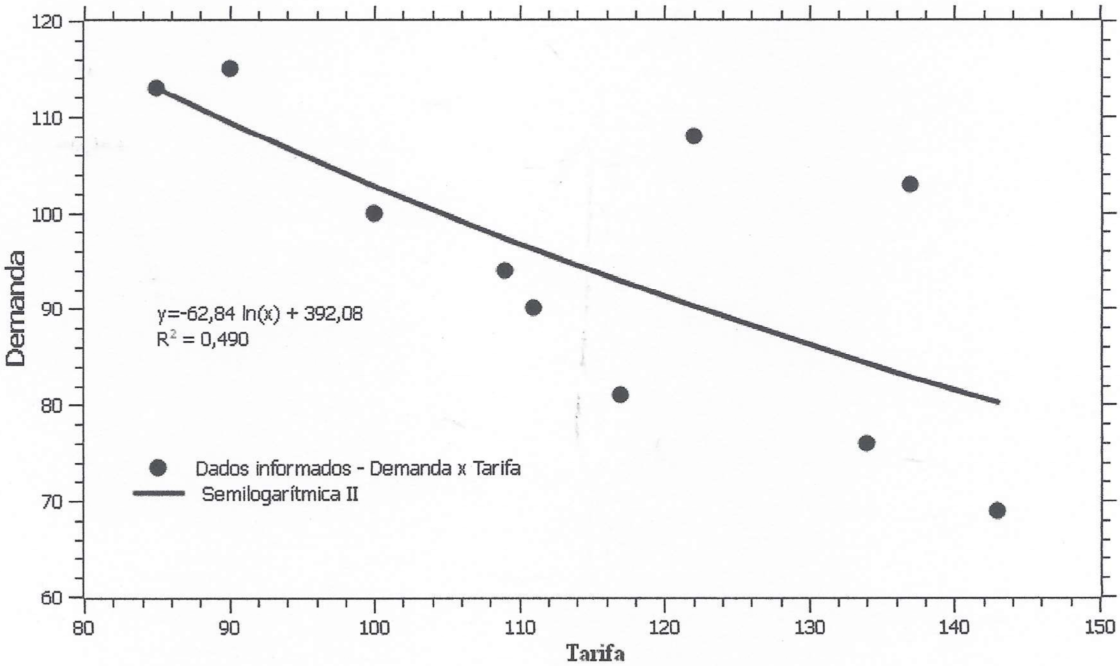


Tabela 16 – Resumo dos Resultados - função Semilogarítmica II

			$F_c = 5{,}318$	
Estatística de regressão			$t_{a,c} = 2{,}306$	
$R^2 =$	0,490		$t_{b,c} = 2{,}306$	
Observações	10			
ANOVA				
	gl	*SQ*	*MQ*	*F*
Regressão	1	1097,787	1097,787	7,683
Resíduo	8	1143,113	142,889	
Total	9	2240,9		
	Coeficientes	*Erro padrão*	*Stat t*	
Interseção	392,081	107,283	3,655	
Ln X	62,835	22,669	-2,772	

Segue tabela com as estatísticas de avaliação a serem consideradas para a escolha da melhor forma funcional.

Tabela 17 – R^2 e Estatística *F* funções: linear, exponencial, potencial e recíproca I

Forma funcional	R^2	***Estatística F***
Linear	0,482	7,436
Exponencial	0,483	7,465
Potencial	0,485	7,520
Semilogarítmica II	0,490	7,683

De acordo com MATOS (2000) como todas as equações têm o mesmo número de variáveis explicativas, foram estimadas com uma amostra de mesmo tamanho (n = 10) e todos os coeficientes tem o sinal correto, a forma Semilogarítmica II apresenta o melhor ajustamento, dado que possui o mais elevado valor de R^2 (0,490) e da estatística *F* (7,683). De qualquer modo, é importante ressaltar que as especificações não diferem significativamente já que os valores dessas estatísticas são virtualmente idênticos. Assim, adotando-se, adicionalmente, o critério da simplicidade, poder-se-ia escolher a forma linear.

Em seguida será realizado o Teste de Hipótese para cada um dos casos. Relembrando:

Regra de decisão para estatística *F*: Se $F_{Calc} \geq F_{Crit}$, então Rejeita H_0

Regra de decisão para estatística t_a: Se $|t_a| \geq |t_{C,a}|$, então Rejeita H_0

Regra de decisão para estatística t_b: Se $|t_b| \geq |t_{C,b}|$, então Rejeita H_0

Segue Tabela 17 que mostra as estatísticas *F* e *t* calculadas e crítico e as respectivas decisões, conforme a função:

Tabela 18 - Estatísticas *F* e *t* funções: linear, exponencial, potencial e recíproca I

Forma funcional	**Calculado**	**Crítico**	**Decisão**
Linear	F = 7,436	F = 5,318	Presença de efeito
	t_a = 6,688	t_a = 2,306	Presença de efeito
	t_b = -2,727	t_b = 2,306	Presença de efeito
Exponencial	F = 7,465	F = 5,318	Presença de efeito
	t_a = 20,109	t_a = 2,306	Presença de efeito
	t_b = -2,732	t_b = 2,306	Presença de efeito
Potencial	F = 7,520	F = 5,318	Presença de efeito
	t_a = 6,571	t_a = 2,306	Presença de efeito
	t_b = -2,742	t_b = 2,306	Presença de efeito

Semilogarítmica II	$F = 7,683$	$F = 5,318$	Presença de efeito
	$t_a = 3,655$	$t_a = 2,306$	Presença de efeito
	$t_b = -2,772$	$t_b = 2,306$	Presença de efeito

Segue abaixo a Análise de Variância com dados reais. A amostra retirada de uma tabela disponível do site do Instituto Brasileiro de Geografia e Estatística (IBGE) informa o rendimento médio mensal de determinada classe de trabalhadores de cinco regiões metropolitanas no período de fevereiro de 2002 a janeiro de 2016. Pela grande quantidade de dados a tabela foi adaptada trabalhando-se apenas com a metrópole de São Paulo no período de 2011 a 2015 totalizando 60 amostras.

De forma análoga ao que foi feito anteriormente, seguem os resultados:

1. Forma Linear: gráfico e quadro de análise (Anova)

Figura 11 - Forma Linear - dados IBGE

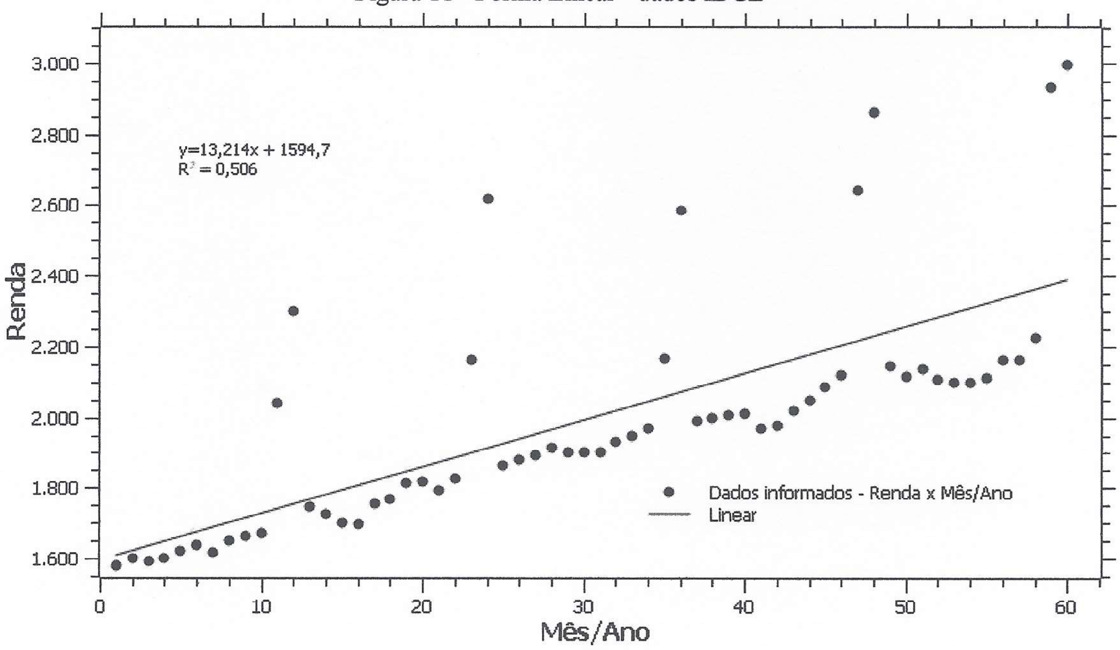


Tabela 19 – Resumo dos Resultados - função linear - dados IBGE

			$F_c = 4{,}007$	
Estatística de regressão			$t_{a,c} = 2{,}002$	
$R^2 =$ 0,506			$t_{b,c} = 2{,}002$	
Observações 60				
ANOVA				
	gl	*SQ*	*MQ*	*F*
Regressão	1	3142130,236	3142130,236	59,453
Resíduo	58	3065348,561	52850,837	
Total	59	6207478,797		
	Coeficientes	*Erro padrão*	*Stat t*	
Interseção	1594,706	60,108	26,531	
X - Mês/Ano	13,214	1,714	7,711	

2. Forma Exponencial: gráfico e quadro de análise (Anova)

Figura 12 - Forma Exponencial - dados IBGE

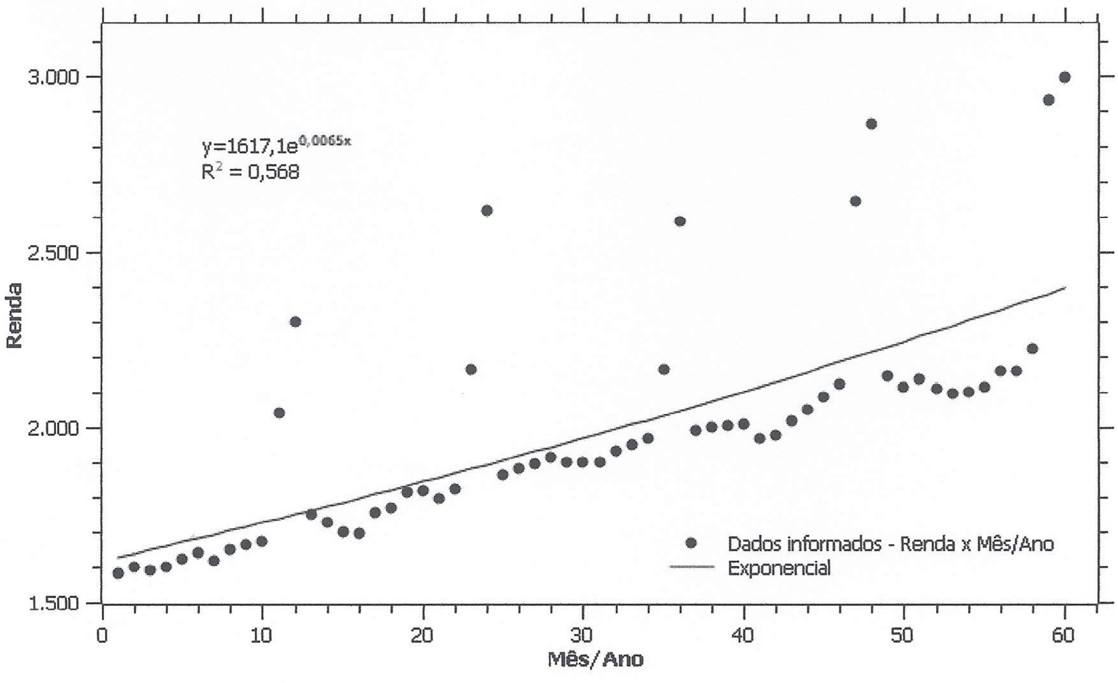


Tabela 20 – Resumo dos Resultados - função exponencial - dados IBGE

			$F_c = 4{,}007$	
Estatística de regressão			$t_{a,c} = 2{,}002$	
$R^2 =$	0,568		$t_{b,c} = 2{,}002$	
Observações	60			
ANOVA				
	gl	*SQ*	*MQ*	*F*
Regressão	1	0,771	0,771	76,198
Resíduo	58	0,587	0,010	
Total	59	1,357		
	Coeficientes	*Erro padrão*	*Stat t*	
Interseção	7,388	0,026	280,975	
X - Mês/Ano	0,007	0,001	8,729	

3. Forma Potencial: gráfico e quadro de análise (Anova)

Figura 13 - Forma Potencial - dados IBGE

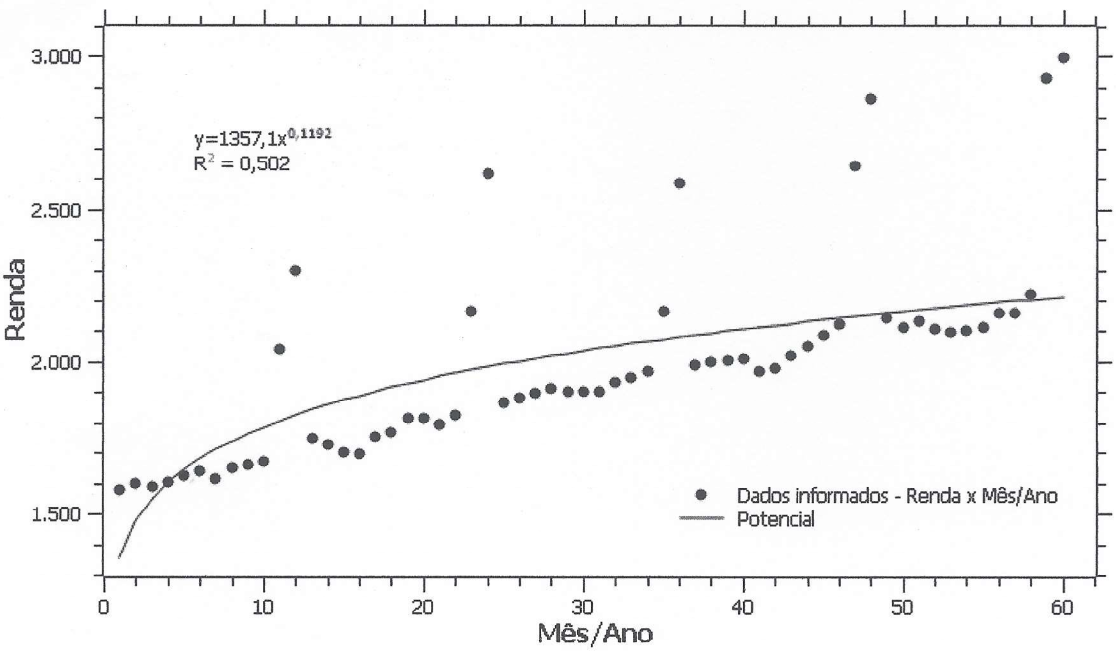


Tabela 21 – Resumo dos Resultados - função potencial - dados IBGE

			$F_c = 4{,}007$	
Estatística de regressão			$t_{a,c} = 2{,}002$	
$R^2 =$	0,502		$t_{b,c} = 2{,}002$	
Observações	60			
ANOVA				
	gl	*SQ*	*MQ*	*F*
Regressão	1	0,681	0,681	58,364
Resíduo	58	0,677	0,012	
Total	59	1,357		
	Coeficientes	*Erro padrão*	*Stat t*	
Interseção	7,213	0,051	141,391	
Ln X	0,119	0,016	7,640	

4. Forma Semilogarítmica II: gráfico e quadro de análise (Anova)

Figura 14 - Forma Semilogarítmica II - dados IBGE

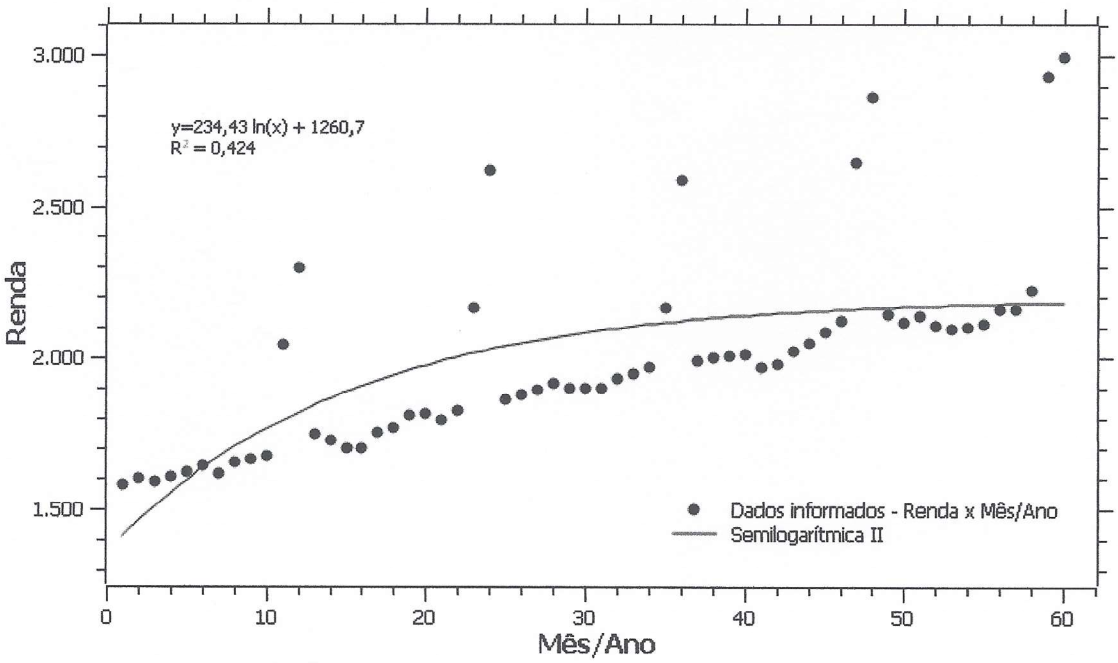


Tabela 22 – Resumo dos Resultados - função Semilogarítmica II - dados IBGE

Estatística de regressão			$F_c = 4{,}007$	
$R^2 =$	0,424		$t_{c,a} = 2{,}002$	
Observações	60		$t_{c,b} = 2{,}002$	
ANOVA				
	gl	*SQ*	*MQ*	*F*
Regressão	1	2631122,518	2631123	42,670
Resíduo	58	3576356,278	61661,32	
Total	59	6207478,797		
	Coeficientes	*Erro padrão*	*Stat t*	
Interseção	1260,729	117,291	10,749	
Ln X	234,431	35,888	6,532	

Segue Tabela 22 com as estatísticas de avaliação a serem consideradas para a escolha da melhor forma funcional.

Tabela 23 - R^2 e Estatística F - dados IBGE

Forma funcional	R^2	***Estatística F***
Linear	0,506	59,453
Exponencial	0,568	76,198
Potencial	0,502	58,364
Semilogarítmica II	0,424	42,670

Como todas as equações têm o mesmo número de variáveis explicativas, foram estimadas com uma amostra de mesmo tamanho (n = 60) e todos os coeficientes tem o sinal correto, a forma Exponencial apresenta o melhor ajustamento, dado que possui o mais elevado valor de R^2 (0,568) e da estatística F (76,198). Em seguida será realizado o Teste de Hipótese para cada um dos casos. Relembrando:

Regra de decisão para estatística F: Se $F_{Calc} \geq F_{Crit}$, então Rejeita H_0

Regra de decisão para estatística t_a: Se $|t_a| \geq |t_{C,a}|$, então Rejeita H_0

Regra de decisão para estatística t_b: Se $|t_b| \geq |t_{C,b}|$, então Rejeita H_0

Segue Tabela 23 que mostra as estatísticas F e t calculadas e crítico e as respectivas decisões, conforme a função:

Tabela 24 - Estatísticas F e t - dados IBGE

Forma funcional	Calculado	Crítico	Decisão
Linear	$F = 59{,}453$	$F = 4{,}007$	Presença de efeito
	$t_a = 7{,}711$	$t_a = 2{,}002$	Presença de efeito
	$t_b = 26{,}531$	$t_b = 2{,}002$	Presença de efeito
Exponencial	$F = 76{,}198$	$F = 4{,}007$	Presença de efeito
	$t_a = 8{,}729$	$t_a = 2{,}002$	Presença de efeito
	$t_b = 280{,}975$	$t_b = 2{,}002$	Presença de efeito
Potencial	$F = 58{,}364$	$F = 4{,}007$	Presença de efeito
	$t_a = 7{,}640$	$t_a = 2{,}002$	Presença de efeito
	$t_b = 141{,}391$	$t_b = 2{,}002$	Presença de efeito
Semilogarítmica II	$F = 42{,}670$	$F = 4{,}007$	Presença de efeito
	$t_a = 10{,}749$	$t_a = 2{,}002$	Presença de efeito
	$t_b = 6{,}532$	$t_b = 2{,}002$	Presença de efeito

Como os valores das estatísticas F e t calculados são maiores que os valores críticos ou valores tabelados segue que o efeito conjunto das variáveis explicativas, no caso apenas uma por ser regressão linear simples e o efeito individual da variável explicativa Mês/Ano e do termo constante exercem influência sobre a variável dependente, pois há presença de efeito.

Um ponto a ser destacado é a pequena diferença de valores de R^2 calculado: 0,506; 0,568; 0,502 e 0,424. É possível especular que os pontos referentes às parcelas de 13° salário podem estar influenciando em seu cálculo. Como os pontos equivalentes a primeira e a segunda parcelas destoam dos demais é possível que haja interferência no cálculo do coeficiente de determinação. Um teste a ser feito seria o de ignorar esses pontos ou dividir o excesso para os meses do ano todo e verificar quais seriam os novos parâmetros e efetuar o cálculo de R^2. Será que o R^2 seria mais próximo de "1" demonstrando que a reta de regressão se ajusta melhor aos dados?

Conclusões

Através dos estudos e análises efetuados neste Trabalho de Conclusão de Curso, verificou-se que o Método dos Mínimos Quadrados (MMQ) é um método eficiente para determinar os parâmetros desconhecidos de uma equação de condições, pois o coeficiente de correlação de Pearson, r, e o coeficiente de determinação, R^2, associados ao MMQ o comprova.

Neste trabalho estudou-se regressão linear e não linear simples com duas variáveis, aplicado à área de Matemática Aplicada, especificamente Econometria, onde se buscou encontrar o melhor ajuste para um conjunto de dados minimizando a soma dos quadrados das diferenças entre o valor estimado e o valor observado.

Ressalta-se a importância de conseguir chegar a uma equação que tratando com dados reais permite que decisões possam ser tomadas ou fenômenos possam ser compreendidos. No caso analisado consumo de energia elétrica residencial por ano em Sorocaba-SP têm-se uma situação importante, onde de acordo com a previsão, segundo a equação encontrada o consumo de energia estava aumentando em um ritmo não ideal e decisões importantes e tempestivas, no caso, devem ser tomadas. O município de Sorocaba, assim como outros, passou por programa governamental onde o consumo de energia foi reduzido. Infere-se que houve acompanhamento através de cálculos estatísticos chegando-se a uma equação estimada. No caso da análise da renda média do trabalhador ao longo dos anos é possível fazer comparativos em relação à inflação e analisar o poder de compra efetivo e a quantidade de capital injetado na economia, por exemplo.

Portanto, conclui-se que a utilização da Regressão Linear e não Linear associada às Estatísticas de Avaliação e Teste de Hipóteses permite avaliar o quão eficiente é um modelo estimado.

Entretanto as possibilidades de estudos e investigações não se esgotam podendo ser ampliadas para outras áreas do conhecimento humano com outras formas funcionais.

Referências Bibliográficas

[1] Consumo de Energia Elétrica, por Setores. Disponível em : <http://www.imp.seade.gov.br/frontend/#/tabelas> Acesso em: 2016-11-23.

[2] Distribuição t-Student. Disponível em: http://www.ibilce.unesp.br/#!/departamentos/cienc-comp-estatistica/docentes/adriana-barbosa-santos/ensino-de-graduacao/> Acesso em: 2017-09-10.

[3] Distribuição F. Disponível em <http://www.ibilce.unesp.br/#!/departamentos/cienc-comp-estatistica/docentes/adriana-barbosa-santos/ensino-de-graduacao/> Acesso em: 2017-09-10.

[4] GUJARATI, D. N., DAWN, C. P. **Econometria Básica**. São Paulo: McGraw Hill, 2000.

[5] HILL, R.C., JUDGE, G.G., GRIFFITHS, W.E. **Econometria**. São Paulo: Saraiva, 2010.

[6] MATOS, O. C. **Econometria Básica: Teoria e Aplicações**. São Paulo: Atlas, 2000.

[7] Renda Média por Metrópole. Disponível em: <http://www.ibge.gov.br/home/estatistica/indicadores/trabalhoerendimento/pme_nova/defaulttab_hist.shtm> Acesso em 2017-05-24

[8] R Development Core Team (2011). *R:* **A language and environment for statistical computing**. R Foundation for Statical Computing, Vienna, Austria. Disponível em <http://www.R-project.org/.

[9] Scidavis. Disponível em: <http://www.scidavis.sourceforce.net>. Acesso em: 24 mai 2017.

[10] Sorocaba é a oitava em consumo de energia elétrica. Disponível em: <http://www.jornalcruzeiro.com.br/materia/741939/sorocaba-e-a-oitava-em-consumo-de-energia-eletrica> Acesso em: 02 dez 2016.

[11] SPIEGEL, M.R., STEPHENS, L.J. **Estatística** – Coleção Schaum. Ed. Bookman, 2009.

[12] STOCK, J.H., WATSON, M.W. **Econometria.** São Paulo, 2004.

[13] VASCONCELLOS, M.A.S, ALVES, D. **Manual de Econometria: Nível intermediário.** São Paulo: Atlas, 2000.

[14] VENEZUELA, A. L. **Modelagem Analítico-Numérica do Escoamento Laminar Convectivo em Tubos Associada à Filtração Tangencial.** Tese (Doutorado em Térmica e Fluidos) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2008. Disponível em: <http://www.teses.usp.br/teses/disponiveis/18/18147/tde-17012011-141704/>. Acesso em: 2016-01-18.

Anexo

Dados do Instituto Brasileiro de Geografia e Estatística (IBGE) adaptados para utilização nesse trabalho. Os dados originais são do período de fevereiro de 2002 a janeiro de 2016. No trabalho foram utilizados os dados da região metropolitana de São Paulo no período de janeiro de 2011 a dezembro de 2015.

Tabela 119 - Rendimento médio nominal do trabalho principal, efetivamente recebido no mês de referência, pelas pessoas de 10 anos ou mais de idade, empregadas com carteira de trabalho assinada do setor privado, no trabalho principal da semana de referência, por regiões metropolitanas, segundo os meses de referência - jan.2011-dez.2015

Meses de referência	Rendimento médio nominal do trabalho principal, efetivamente recebido no mês de referência, pelas pessoas de 10 anos ou mais de idade, empregadas com carteira de trabalho assinada do setor privado, no trabalho principal da semana de referência						
	Total	Regiões metropolitanas					
		Recife	Salvador	Belo Horizonte	Rio de Janeiro	São Paulo	Porto Alegre
	Estimativas (Em R$)						
Janeiro 11	1425,6	1027,2	1126,8	1188,3	1470,8	1.580,00	1286,1
Fevereiro	1434,7	1031,1	1117,0	1216,8	1466,1	1.599,80	1260,6
Março	1440,3	1005,4	1178,5	1243,6	1489,8	1.590,50	1260,9
Abril	1457,0	1004,2	1193,5	1266,6	1528,4	1.602,10	1255,1
Maio	1462,0	963,1	1196,0	1325,4	1476,0	1.621,50	1294,0
Junho	1482,0	1040,0	1214,4	1325,2	1487,8	1.638,80	1310,2
Julho	1466,9	1052,4	1208,3	1292,1	1495,5	1.616,10	1282,1
Agosto	1480,6	1001,9	1247,2	1259,1	1492,2	1.649,10	1302,0
Setembro	1486,1	1094,4	1272,3	1278,0	1446,7	1.661,50	1298,2
Outubro	1499,0	1099,8	1260,9	1266,1	1488,2	1.671,50	1303,1
Novembro	1740,6	1109,1	1311,1	1387,9	1685,2	2.042,00	1404,6
Dezembro	2094,1	1456,1	1901,6	1784,4	2089,4	2.299,20	1933,1
Janeiro 12	1575,7	1124,3	1342,6	1359,9	1589,9	1.747,90	1354,3
Fevereiro	1576,3	1131,3	1366,2	1386,0	1622,3	1.726,60	1349,7
Março	1553,2	1061,1	1331,4	1394,5	1597,8	1.700,90	1359,9
Abril	1557,8	1126,6	1304,1	1401,2	1601,3	1.696,90	1387,9
Maio	1589,2	1116,2	1323,6	1495,2	1575,4	1.753,30	1365,8
Junho	1596,4	1136,5	1307,6	1436,1	1590,4	1.767,70	1394,7
Julho	1626,3	1151,3	1338,6	1417,2	1621,8	1.811,90	1421,0
Agosto	1639,4	1141,9	1376,2	1433,4	1649,6	1.816,50	1439,4
Setembro	1640,2	1156,9	1325,2	1459,3	1680,8	1.793,40	1487,2
Outubro	1652,6	1177,2	1328,9	1452,5	1687,7	1.824,80	1441,1
Novembro	1873,6	1275,2	1385,9	1508,2	1862,7	2.163,60	1532,2
Dezembro	2340,3	1553,3	1897,2	1976,4	2350,1	2.616,50	2120,7
Janeiro 13	1695,9	1218,6	1361,8	1504,5	1746,5	1.863,20	1485,8
Fevereiro	1706,6	1224,7	1343,1	1495,0	1783,1	1.880,70	1444,9
Março	1702,0	1236,1	1315,6	1515,2	1730,8	1.894,90	1455,0
Abril	1721,1	1199,6	1320,1	1500,8	1776,4	1.914,00	1477,0
Maio	1724,0	1253,9	1360,8	1484,3	1764,7	1.899,80	1568,5
Junho	1723,3	1156,1	1345,2	1542,0	1760,4	1.900,00	1576,3
Julho	1747,2	1209,6	1356,3	1563,8	1823,5	1.901,00	1600,7
Agosto	1760,6	1234,3	1369,2	1545,5	1818,7	1.931,30	1583,8
Setembro	1769,2	1236,2	1370,5	1539,7	1812,4	1.948,00	1624,1
Outubro	1796,4	1261,6	1315,1	1575,7	1866,0	1.967,70	1688,8
Novembro	1953,6	1288,5	1383,7	1618,5	2028,1	2.164,40	1920,2

Dezembro	2372,1	1990,4	1730,9	1791,9	2492,1	2.586,40	2343,7
Janeiro 14	1840,3	1274,8	1446,2	1608,9	1970,7	1.991,20	1682,9
Fevereiro	1857,7	1331,3	1432,3	1599,1	2034,9	1.998,20	1662,9
Março	1856,5	1365,8	1428,1	1609,4	2006,9	2.006,70	1637,1
Abril	1881,2	1364,4	1449,2	1611,3	2081,8	2.010,00	1693,7
Maio	1848,3	1377,1	1420,9	1598,3	2042,9	1.968,60	1670,2
Junho	1862,5	1359,9	1424,8	1604,1	2071,0	1.978,50	1713,1
Julho	1886,8	1383,8	1420,7	1661,4	2055,6	2.019,90	1739,0
Agosto	1897,2	1397,7	1389,2	1654,1	2009,1	2.050,10	1784,7
Setembro	1929,1	1362,2	1446,1	1700,6	2063,5	2.086,30	1790,3
Outubro	1941,3	1361,9	1468,0	1696,5	2028,2	2.121,10	1792,2
Novembro	2257,3	1639,8	1577,4	1667,2	2222,2	2.644,80	1913,4
Dezembro	2574,1	2114,6	1826,0	1905,5	2729,9	2.862,60	2332,7
Janeiro 15	1986,6	1478,5	1550,9	1728,5	2117,8	2.144,10	1792,5
Fevereiro	1970,3	1446,2	1458,5	1692,0	2154,2	2.115,00	1795,8
Março	1964,4	1373,7	1423,1	1686,5	2123,4	2.135,60	1756,5
Abril	1960,8	1382,4	1469,0	1685,1	2130,5	2.107,30	1804,9
Maio	1979,4	1411,0	1508,3	1697,9	2192,0	2.096,70	1872,9
Junho	1978,6	1440,1	1528,7	1718,4	2166,0	2.100,50	1882,4
Julho	2008,5	1403,1	1574,7	1724,9	2262,5	2.112,10	1887,2
Agosto	2005,9	1431,9	1526,9	1743,8	2145,8	2.161,00	1870,9
Setembro	1992,4	1404,5	1542,7	1702,2	2123,3	2.160,30	1827,0
Outubro	2032,8	1422,2	1601,3	1728,4	2111,4	2.222,70	1889,0
Novembro	2.407,3	1.617,0	1.615,9	1.748,8	2.211,3	2.932,20	1.931,3
Dezembro	2.701,6	2.157,1	1.862,7	2.339,4	2.640,8	2.995,10	2.697,2

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Mensal de Emprego jan.2011-dez.2015.
Nota: Exclusive trabalhadores domésticos.

Segue tabela estatística Distribuição *t-Student* ao nível de 5% de probabilidade.

Tabela 25 - Distribuição *t-Student*

Tabela 5. Limites unilaterais da distribuição F de Fisher-Snedecor ao nível de 5% de probabilidade.

GL	V1																			
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.0	243.9	244.7	245.4	245.9	248.0	251.1	252.2	253.3	253.8
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396	19.405	19.412	19.419	19.424	19.429	19.446	19.471	19.479	19.487	19.492
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.763	8.745	8.729	8.715	8.703	8.660	8.594	8.572	8.549	8.538
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.936	5.912	5.891	5.873	5.858	5.803	5.717	5.688	5.658	5.643
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.704	4.678	4.655	4.636	4.619	4.558	4.464	4.431	4.398	4.382
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.976	3.956	3.938	3.874	3.774	3.740	3.705	3.687
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.550	3.529	3.511	3.445	3.340	3.304	3.267	3.249
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.259	3.237	3.218	3.150	3.043	3.005	2.967	2.947
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.048	3.025	3.006	2.936	2.826	2.787	2.748	2.727
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.887	2.865	2.845	2.774	2.661	2.621	2.580	2.559
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.761	2.739	2.719	2.646	2.531	2.490	2.448	2.426
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.660	2.637	2.617	2.544	2.426	2.384	2.341	2.319
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.577	2.554	2.533	2.459	2.339	2.297	2.252	2.230
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.507	2.484	2.463	2.388	2.266	2.223	2.178	2.155
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.448	2.424	2.403	2.328	2.204	2.160	2.114	2.090
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.456	2.425	2.397	2.373	2.352	2.276	2.151	2.106	2.059	2.035
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.413	2.381	2.353	2.329	2.308	2.230	2.104	2.058	2.011	1.986
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.374	2.342	2.314	2.290	2.269	2.191	2.063	2.017	1.968	1.943
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.340	2.308	2.280	2.256	2.234	2.155	2.026	1.980	1.930	1.905
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.310	2.278	2.250	2.225	2.203	2.124	1.994	1.946	1.896	1.870
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.283	2.250	2.222	2.197	2.176	2.096	1.965	1.916	1.866	1.839
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.259	2.226	2.198	2.173	2.151	2.071	1.938	1.889	1.838	1.811
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.236	2.204	2.175	2.150	2.128	2.048	1.914	1.865	1.813	1.785
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.216	2.183	2.155	2.130	2.108	2.027	1.892	1.842	1.790	1.762
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.198	2.165	2.136	2.111	2.089	2.007	1.872	1.822	1.768	1.740
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220	2.181	2.148	2.119	2.094	2.072	1.990	1.853	1.803	1.749	1.720
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204	2.166	2.132	2.103	2.078	2.056	1.974	1.836	1.785	1.731	1.702
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.151	2.118	2.089	2.064	2.041	1.959	1.820	1.769	1.714	1.685
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.138	2.104	2.075	2.050	2.027	1.945	1.806	1.754	1.698	1.669
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.126	2.092	2.063	2.037	2.015	1.932	1.792	1.740	1.683	1.654
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.038	2.003	1.974	1.948	1.924	1.839	1.693	1.637	1.577	1.544
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026	1.986	1.952	1.921	1.895	1.871	1.784	1.634	1.576	1.511	1.476
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.952	1.917	1.887	1.860	1.836	1.748	1.594	1.534	1.467	1.430
80	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951	1.910	1.875	1.845	1.817	1.793	1.703	1.545	1.482	1.411	1.370
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.886	1.850	1.819	1.792	1.768	1.676	1.515	1.450	1.376	1.333
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910	1.869	1.834	1.803	1.775	1.750	1.659	1.495	1.429	1.352	1.307
240	3.881	3.033	2.642	2.409	2.252	2.136	2.048	1.977	1.919	1.870	1.829	1.793	1.761	1.733	1.708	1.614	1.445	1.375	1.290	1.237

Fonte: Unesp

Segue tabela estatística Distribuição *F*. Considerando ao nível de 5% de probabilidade, foi usada a coluna 0,025% de cada lado.

Tabela 26 - Distribuição F

Distribuição t-Student

Valores associados à cauda direita

	α									
GL	**0,4**	**0,25**	**0,2**	**0,15**	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**	**0,0005**
1	0,325	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,289	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,277	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,271	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,267	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,265	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,263	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,262	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,261	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,260	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,260	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,259	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,259	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,258	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,258	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,258	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,257	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,257	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,257	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,257	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,257	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,256	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,256	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,256	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,256	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,256	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,256	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,689
28	0,256	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,256	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,660
30	0,256	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,255	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
50	0,255	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678	3,496
60	0,254	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,460
70	0,254	0,678	0,847	1,044	1,294	1,667	1,994	2,381	2,648	3,435
80	0,254	0,678	0,846	1,043	1,292	1,664	1,990	2,374	2,639	3,416
90	0,254	0,677	0,846	1,042	1,291	1,662	1,987	2,368	2,632	3,402
100	0,254	0,677	0,845	1,042	1,290	1,660	1,984	2,364	2,626	3,390
120	0,254	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
oo	0,253	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,290

Fonte: Unesp